Technological Change and Immigration Policy

Gerardo Gomez-Ruano

June 2011

Technological Change and Immigration Policy

Gerardo Gomez-Ruano [*†]

June 2011

Abstract

We propose a dynamic general equilibrium model to address the effects of technological progress on immigrant skill composition. Our results from this positive model suggest that neutral and skill-biased technological change imply essentially different immigration policies. On the one hand, skill-neutral change implies an immigrant skill distribution that is dominated by the native skill distribution; on the other hand, skill-biased change implies an immigrant over-representation at the top and bottom of the skill distribution. This result is interesting because of its unexpected nature. It implies that if technology changes as it has in the last decades and education has an increasing cost, then it is optimal to allow some low-skill immigration along with high-skill immigration. We show consistency of our model's predictions with data from the United States and Canada.

Keywords: immigration, education, complementarity, technological change, mobility, skill bias.

[*]Department of Economics, Universidad Iberoamericana, Mexico. Email: ggomezr@alum.bu.edu

[†]The author is grateful for the continuing advice and support from Kevin Lang, Bart Lipman and Zvika Neeman. He also benefited from talks with Gordon Hanson, Andrew Newman, Dilip Mookherjee, Michael Manove, Joseph Altonji, Robert Margo, Jeffrey Ely, Daniele Paserman and comments from other participants at the Boston University workshops as well as from the encouragement from Jörgen Weibull.

"The greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgement with which it is anywhere directed, or applied, seem to have been the effects of the division of labour."

Adam Smith

1 Introduction

While all existing theories on the subject suggest that immigration is either high-skilled or low-skilled, existing data for at least two of the traditional destination countries (Canada and USA) shows that immigration occurs at both ends of the skill distribution, as Figure 1 shows.[1] In addition to providing the evidence, this paper introduces a model where skill-bias and strong risk aversion may induce the aforementioned pattern.

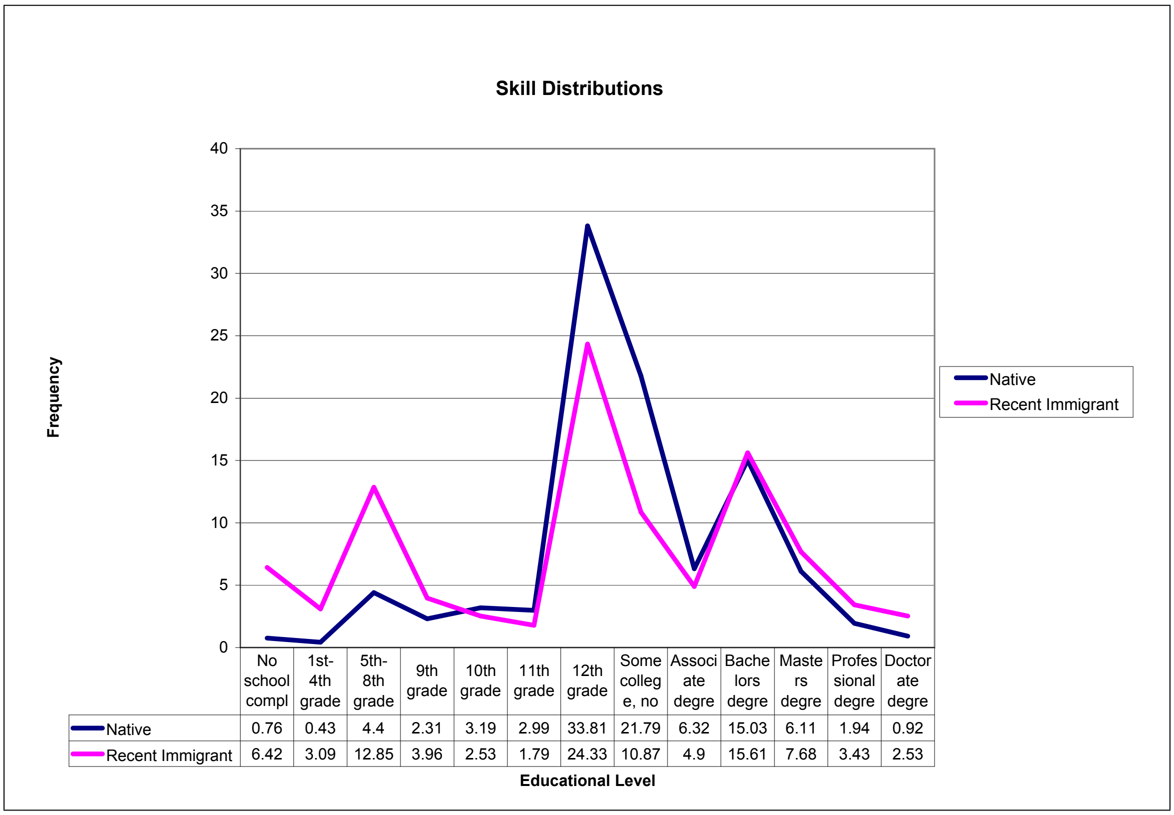
In what follows, this section explains the two main strands of immigration theories, which are sometimes referred to as supply-driven and demand-driven theories. Consider first the well-known Borjas-Roy theory, which is a prototypical supply-driven theory.[2] In this static framework, labor units are homogeneous and therefore perfectly substitutable, with the logarithm of units provided (and hence the logarithm of wage) being proportional to skill plus ability, which is normally distributed. The different levels of inequality across countries of origin then induces emigration of either the low-skilled (given high inequality) or the high-skilled (given low inequality). It has been shown by Chiquiar and Hanson (2005) that this characterization is inaccurate. They show this in the case of Mexican immigrants in the USA, which is non-trivial in size; the great majority is drawn from the middle of the Mexican income distribution. Moreover, the Borjas-Roy model, as all of the supply-driven models (by definition), assumes there are no immigration restrictions enforced by the governments of destination countries. This goes against widely recognized facts, especially among traditional destination countries.[3]

There has been a surge of demand-driven models in response to this last evident flaw. The political economy models by Benhabib (1996) and Ortega (2005) are probably the most representative of this strand. The latter concludes that at one point in time the incoming immigrant's skill composition is either dominantly high-skilled or dominantly

[1]Figure 1 displays the distributions of school attainment for natives (individuals born on US soil) and recent immigrants (individuals who were born outside the US but entered within the five years prior to the 2000 Census). All individuals are 30 years or older.

[2]Borjas (1987).

[3]The USA, for example, according to its official immigration policy, spends a non-trivial amount of resources on border enforcement in addition to favoring high-skilled immigration. Canada and Australia both have official point-systems for immigrants as well. See OECD (2001).



	No school compl	1st-4th grade	5th-8th grade	9th grade	10th grade	11th grade	12th grade	Some college, no	Associate degre	Bachelors degre	Masters degre	Professional degre	Doctorate degre
Native	0.76	0.43	4.4	2.31	3.19	2.99	33.81	21.79	6.32	15.03	6.11	1.94	0.92
Recent Immigrant	6.42	3.09	12.85	3.96	2.53	1.79	24.33	10.87	4.9	15.61	7.68	3.43	2.53

Figure 1: Skill Distributions, 30+ years old individuals, 2000 Census USA

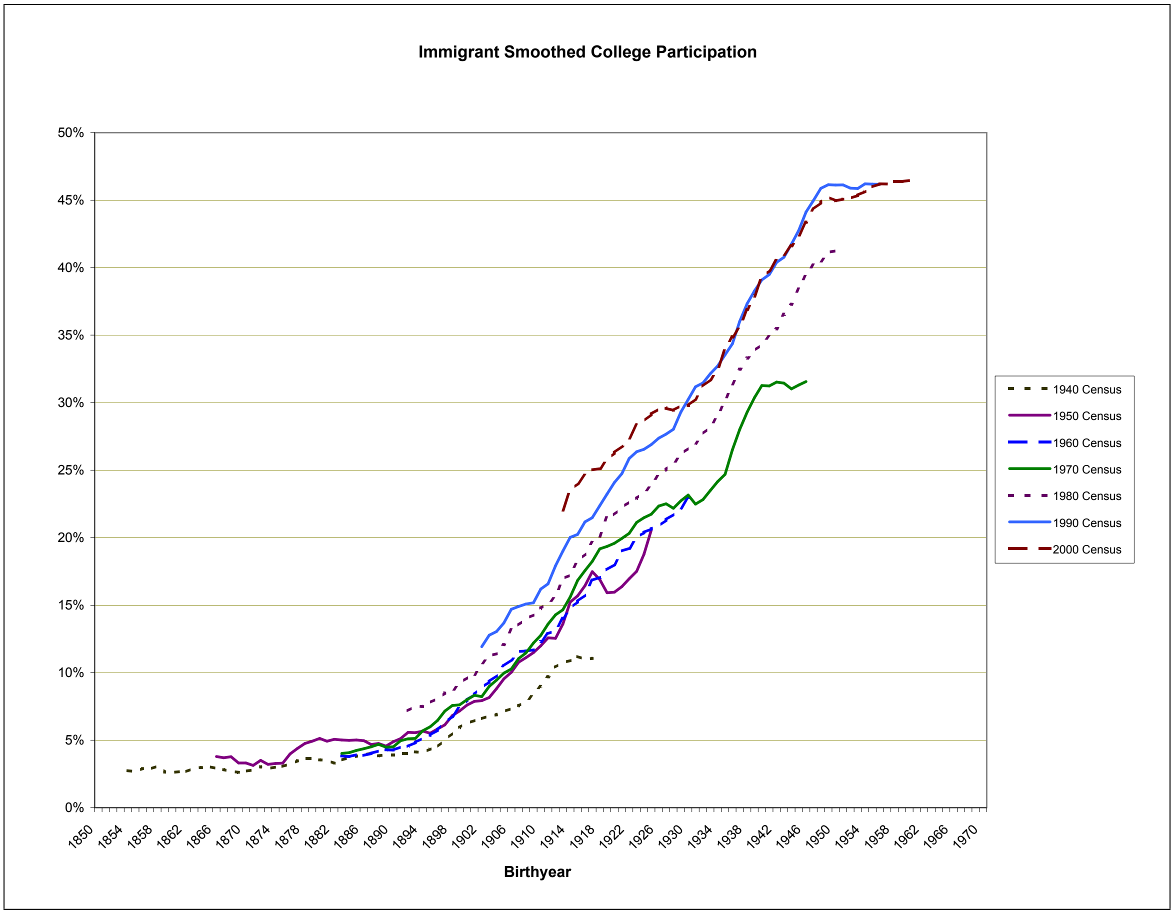


Figure 2: Immigrant College Participation, 1940-2000 Censuses

low-skilled. Although the latter case may be accurate under some raw measures of skill,[4] there are two other major problems with this approach that have been pointed out by Gomez-Ruano (2007). First, it contradicts the facts when trying to explain the actual, increasing college participation of immigrants observed in the data (see Figure 2).[5] Second, it remains silent with respect to the overrepresentation of immigrants at both ends of the skill distribution shown in Figure 1.

[4]Specifically under aggregation into two skills: low skilled being those with at most high-school, and high-skilled being the better educated rest.

[5]Figure 2 displays the College Participation (i.e., percentage of individuals who finished College) for each cohort according to each respective Census. The series were smoothed with a moving average for the sake of presentation.

The rest of the paper is organized as follows: Section 2 covers the theoretical model and results. Section 3 covers the empirical evidence. Section 4 provides concluding remarks. The reader might find our *quick reference for notation* in Appendix A on page 37 useful, though we have tried to keep the technical notation (in the main text) at a reasonable level.

2 Theoretical findings

2.1 Preliminaries

Our point of departure is a dynamic general equilibrium model where an exogenous skill-specific technological process determines the efficiency of each of the inputs from an otherwise fairly typical production function, thus affecting their marginal product.

Our social setup is a continuum of one-adult-one-child families, where the adult is the choice maker who divides his labor earnings between (family) consumption and schooling expenditures, the latter being the only (imperfect) way to transfer wealth by affecting the earnings prospects of the child.

We assume there is a finite number of skills (and hence earnings levels); however our model still allows for an i.i.d. ability interpretation and assumes a continuous schooling-expenditures choice as well.

We do impose one key assumption, namely that skills are strictly ordered,[6] slightly complement each other (i.e., there is a finite elasticity of substitution between them), and that workers are able to perform all jobs that require an equal or lower skill than their own. In other words, skills are *downward substitutable and upward complementary*, an assumption that is consistent with a one-dimensional specialization of labor.[7]

Another key assumption we make is that labor is the only production factor.[8] Since our interest is on skill composition and not on size of immigrant flows, this is a natural and greatly simplifying assumption of no further consequence for our results, which are of qualitative character. As a result of this, however, family income derives exclusively from workers' wages and hence wealth heterogeneity *within* skills is avoided.

It therefore follows that workers, whether immigrant or native, are fully characterized by their skill level. Moreover, their only choice is how much of their earnings to spend on consumption (and hence how much to spend on their child's education) given their altruistic/filial preferences and their expectations of the future.

Keep in mind that in our competitive general-equilibrium approach, tomorrow's wages will be affected by today's schooling-expenditure decisions of the entire workforce, and these, in turn, will be affected by expectations of tomorrow's wages for the entire workforce as well.

[6]Let $\mathcal{M}$ be the set of skills, then for all $a, b \in \mathcal{M}$ with $a \neq b$: $a > b$ or $b > a$. For our purposes we use $\mathcal{M} = \{1, 2, ..., m\}$, with $m > 2$.

[7]In real life, specialization is rarely one-dimensional, thus making an ordering of skills according to this type of "downward substitution" at first sight very unrealistic. A more careful analysis reveals that this is not as big of a limitation.

[8]More specifically, workers' skills are the only production factors.

So far we have described all but the immigration process. Let us now turn to this. We assume that immigration policy is a centralized choice made at the beginning of every period (generation) by some Social Planner (D) that maximizes the sum of all natives' welfare. The consequences from this policy-choice mechanism are not trivial since, as is the case in most countries, children of immigrants are born natives[9] and thus the objective function of such a Social Planner will generally change every period. To make this point as stark as possible, we assume that every generation (time period) has a different Social Planner, i.e., we assume there is a sequence of Social Planners ($\{D_t\}_{t=0}^{\infty}$), one for every period, each of them choosing an immigration policy in order to maximize the added welfare of their native contemporaries. This is a simple device for capturing the voting by natives (and *not* by immigrants).

Thus in a typical time period, the contemporary Social Planner first chooses an immigration policy. Second, production takes place with the whole workforce (natives and immigrants) and receives its payment. And finally, workers spend their earnings on consumption and training/education for their offspring.

Throughout the whole process, decision makers are aware of the structure of the entire system.

We demonstrate that under suitable conditions the infinite sequential problem (SP) satisfies a recursive representation, i.e., a functional equation system (FES). Such representation allows us to further characterize the equilibrium policies for different technological processes, thus providing testable implications. We then show in Section 3 data from the USA and Canada, relate these observed implications with the conditions needed by the model to produce such patterns, and estimate the parameters in question. It is through this exercise that we find the consistency from the model/assumptions with the data.

In this section we first present the model. Next we proceed to define and prove both the existence and uniqueness of equilibrium. We then show that the sequential problem satisfies a functional equation system, a fact that helps us characterize the equilibrium immigration policy under neutral and skill-biased technological change, thus deriving testable implications.

2.2 The Model

We will now present the different elements of our model. First we will introduce preferences; second, we will describe production, followed by schooling; and finally, the tran-

[9] We take this fact as given although one could imagine that citizenship is granted to preserve peace and stability, along the same lines of Acemoglu and Robinson (2000)'s explanation for the extension of franchise in the west.

sition functions (also called Laws of Motion)—one for technology, one for the native distribution, and one for a typical dynasty's skill.[10] Recall that in a typical time period, first the contemporaneous Social Planner chooses an immigration policy, then all workers (native and immigrant) perform their jobs and receive their wages, and finally they divide their income between consumption and schooling for their child.

2.2.1 Social Structure and Preferences

Workers are the main actors in our model. Each represents and makes choices for a simple family, consisting of an offspring and an adult (the worker himself).

Let time be $t = 0, 1, 2, ...$ At any period t there is a set of families/workers $\mathcal{L}_t \subseteq \mathbb{R}$ with positive measure (some native, some immigrant) indexed by j. Workers are distinguished by their skill (also called type). There are m strictly ordered skills[11] and we denote the time-invariant set of skills by $\mathcal{M} = \{1, 2, ..., m\}$.

Notation 2.1 *In what follows, depending on the context, we may use up to three subindexes: $j \in \mathcal{L}$ for workers/families, $i \in \mathcal{M}$ for skills/types, and $t \in \{0, 1, ...\}$ for time. We will always use them in the same order although not always display them all. Thus $\mathcal{L}$ is the set of workers, $\mathcal{L}_i$ is the set of workers with skill i, and $\mathcal{L}_{i,t}$ is the set of workers of type i at time t. Similarly, c is consumption, c_i is consumption of a family whose head is of type i, $c_{i,t}$ is consumption of a worker (we use family and worker interchangeably) with skill i at time t and $c_{j,i,t}$ is consumption of worker j (who is of type i) at time t. Finally, whenever time or skill is not stated (as in $\mathcal{L}, c_i, c_t, s_j$,etc.) the paper refers to the contextual time period and/or skill; if this context is absent what is meant is "for some period t and/or for some skill i."*

All workers at any time t have the same preferences over their dynasty's consumption stream represented by the following *Total Utility* function U:

$$U_{jt}(\{c_{js}\}_{s=0}^{\infty}) = U\left(\{c_{js}\}_{s=t}^{\infty}\right) = E_t\left[\sum_{s=t}^{\infty} \beta^{s-t} u(c_{js})\right] \quad \forall j, \forall t \tag{1}$$

where the consumption stream $\{c_{js}\}_{s=t}^{\infty}$ is a stochastic process, E_t is the expectation operator given information available at time t, β is the discount factor, and $u(\cdot)$ is the *Instant Utility* function.

[10] A dynasty is the sequence of workers/families with the same index.

[11] See footnote 6.

Assumption 1 *The time-invariant Instant Utility function* $u : \mathbb{R}_{++} \longrightarrow \mathbb{R}$ *is given by*

$$u(x) = \begin{cases} \frac{b_1}{1-b_2} x^{1-b_2} & \text{if } \; b_2 \neq 1 \\ b_1 \ln x & \text{if } \; b_2 = 1 \end{cases} \tag{2}$$

where $b_1 > 0$ *and* $b_2 \geq 0$.[12]

It follows from Assumption 1 that $u(\cdot)$ is twice continuously-differentiable as well.

2.2.2 Production and Immigration

Recall the set of workers is $\mathcal{L} \subseteq \mathbb{R}$, and let us define:

1. the (total) workforce L as the measure of the set $\mathcal{L}$;
2. the set of workers with skill i, $\mathcal{L}_i$ as the subset of workers in $\mathcal{L}$ with skill/ of type i; and
3. the workforce with skill i, L_i, as the measure of the set $\mathcal{L}_i$.

The set of workers $\mathcal{L}$ is partitioned into the set of native workers $\mathcal{N}$ and the set of immigrant workers $\mathcal{I}$. Analogous definitions like the above 1-3 apply for $\mathcal{N}$ and $\mathcal{L}$. Naturally $L_i = N_i + I_i$ and $L = N + I = \sum_{i \in \mathcal{M}} L_i = \sum_{i \in \mathcal{M}} N_i + \sum_{i \in \mathcal{M}} I_i$.

Let us now specify the production function $F(\cdot)$.

Assumption 2 *The time-invariant production function* $F : \mathbb{R}^m_+ \longrightarrow \mathbb{R}_+$ *is twice continuously-differentiable, increasing, (jointly) homogeneous of degree one in its* m *arguments, has an elasticity of substitution* $\varepsilon_{p,q} \in (1, \infty)$ *for all* $p, q \in \mathcal{M}$ *with* $p \neq q$*, and its first derivatives are equalized at some vector* $\mathbf{x} \in \mathbb{R}^m_{++}$.

Assumption 2 remarks that the only inputs for production are the m labor types/skills; thus, capital is not an input, as we previously indicated. The rest of the assumption is standard except for the restriction on the elasticity of substitution and the equal marginal products for some strictly positive vector.

[12] Where b_2 is known in the macroeconomic literature as "the inverse of the instantaneous elasticity of intertemporal substitution," while in the microeconomic/financial literature it is known as "the risk aversion coefficient;" see Barro and Sala-i-Martin (2001) for the former, and Mas-Collel et al. (1995) for the latter.

Let $(a_1, a_2, ..., a_m) \in \mathbb{R}^m_{++}$ be the efficiency vector and let the *effective* labor supply of type i, $\hat{L}^s_i$, be given by the linear transformation $\hat{L}^s_i = a_i L^s_i$. Then, the national product Y given inputs $(\hat{L}^s_1, \hat{L}^s_2, ..., \hat{L}^s_m)$ would be given by

$$Y = F(\hat{L}^s_1, \hat{L}^s_2, ..., \hat{L}^s_m) \tag{3}$$

It is possible to show that an increase in the efficiency units of skill i, i.e., an increase in a_i, will increase $i's$ wage, under the same labor supply, only if the elasticity of substitution between i and the skill h is greater than one for distinct skills i, h. Hence the condition in Assumption 2 that was left unexplained before.

Notice that equation 3 hides the technological state, is long, and impractical; we will work with a simpler notation. Since wages are homogeneous of degree zero in labor, the relative supplies are a sufficient statistic to calculate them. Assume for the moment that $L^s_i = L_i \ \forall i$, and let $\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \vdots \\ \lambda_m \end{bmatrix} = \begin{bmatrix} L_1 \\ \vdots \\ \vdots \\ L_m \end{bmatrix} L^{-1}$ and similarly $\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \vdots \\ \vdots \\ \eta_m \end{bmatrix} = \begin{bmatrix} N_1 \\ \vdots \\ \vdots \\ N_m \end{bmatrix} N^{-1}$ and $\boldsymbol{\iota} = \begin{bmatrix} \iota_1 \\ \vdots \\ \vdots \\ \iota_m \end{bmatrix} = \begin{bmatrix} I_1 \\ \vdots \\ \vdots \\ I_m \end{bmatrix} I^{-1}$.

We will call $\boldsymbol{\lambda} \in \Delta^{m-1}$ the *labor distribution,* $\boldsymbol{\eta} \in \Delta^{m-1}$ the *native distribution*, and $\boldsymbol{\iota} \in \Delta^{m-1}$ the *immigrant distribution*;[13] $\hat{\boldsymbol{\lambda}}$ is constructed following the same logic. Let $\mathbb{R}^{d(m)}$ denote the *set of diagonal matrices* of size $m \times m$—that is, let $\mathbb{R}^{d(m)} = \{\mathbf{G} \in \mathbb{R}^{m \times m} | \, i \neq j \implies g_{ij} = 0\}$—and define $\mathbb{R}^{d(m)}_{++}$ analogously. We will write the technological state $\mathbf{A}$ as a diagonal matrix rather than a column vector for notational convenience, so from now on $\mathbf{A} \equiv diag(a_1, a_2, ..., a_m) \in \mathbb{R}^{d(m)}_{++}$, therefore allowing us to write the per-worker output as:

$$\frac{Y}{L} = \frac{F(\hat{L}_1, \hat{L}_2, ..., \hat{L}_m)}{L} = \frac{F(\mathbf{A}(L_1, L_2, ..., L_m))}{L} = F(\mathbf{A}\boldsymbol{\lambda}) = F(\hat{\boldsymbol{\lambda}}) \tag{4}$$

We will use $F(\mathbf{A}\boldsymbol{\lambda})$ most of the time for it displays our two key variables, technology and worker's skill composition, in a short form.

[13] where $\Delta^f = \{\mathbf{x} \in [0,1]^{f+1} | \sum_{i=1}^{f+1} x_i = 1\}$ is called the f-dimensional unit simplex.

Consequently wages for workers of type i are given by[14]

$$w_i = \frac{\partial F(\mathbf{A}\boldsymbol{\lambda})}{\partial \lambda_i} \equiv F_i(\mathbf{A}\boldsymbol{\lambda}) \tag{5}$$

In the last paragraphs we assumed workforce L_i was equal to supply L_i^s. This is true as long as some restrictions—in the following paragraphs—are satisfied.

Assume there is an *outside/international, time-invariant option* $\mathbf{d} \in \mathbb{R}^m_{++}$ such that if the domestic wage $w_i < d_i$ natives of type i will emigrate, and if $w_i > d_i$ the supply of (potential) immigrants of type i will be unlimited.[15] Since there are no restrictions on emigration, but maybe on immigration, for each skill level the wage must be at least equal to the outside option. Formally,

Definition 1 *The labor distributions satisfying outside option restrictions are given by the set* $\mathcal{R}_1(\mathbf{A}, \mathbf{d}) \equiv \{\mathbf{x} \in \mathbb{R}^m_+ | \nabla_{\mathbf{x}} F(\mathbf{A}\mathbf{x}) \geq \mathbf{d}\} \cap \Delta^{m-1}$

It is possible to show that this set is convex in addition to compact.

Next, the *downward-substitutability* assumption on skills implies that wages will be non-decreasing in skill since more skilled workers can always fill lower-skill jobs and will choose to do so if they are better paid. Formally,

Definition 2 *The labor distributions satisfying market-clearing conditions are given by the set* $\mathcal{R}_2(\mathbf{A}) \equiv \{\mathbf{x} \in \mathbb{R}^m_+ | F_i(\mathbf{A}\mathbf{x}) \geq F_{i-1}(\mathbf{A}\mathbf{x}) \quad \forall i > 1\} \cap \Delta^{m-1}$

Again, this set is compact, and moreover:

Proposition 1 $\mathcal{R}_2(\mathbf{A})$ *is convex.*

Proof: In Appendix B.1

Consequently, the set of feasible labor distributions $\mathcal{F}(\mathbf{A}, \mathbf{d}) \equiv \mathcal{R}_1(\mathbf{A}, \mathbf{d}) \cap \mathcal{R}_2(\mathbf{A})$ is a convex and compact set. Furthermore, the supply $\boldsymbol{\lambda}^s$ is equal[16] to $\boldsymbol{\lambda}$ as long as $\boldsymbol{\lambda} \in \mathcal{F}(\mathbf{A}, \mathbf{d})$.

To guarantee the existence of a $\boldsymbol{\lambda}$ we make the following:

---

[14]Notice $F_i(\mathbf{A}\boldsymbol{\lambda}) \neq \frac{\partial F(\mathbf{A}\boldsymbol{\lambda})}{\partial (a_i \lambda_i)}$; that is, we take the explicit technological parameter $\mathbf{A}$ as part of the function and not as an argument.

[15]w_i is a sufficient statistic for the worker's Instant Utility but not for the Total Utility. It can be shown, however, that this simplification is innocuous for our purposes.

[16]Technically it is equal to the supply function $\boldsymbol{\lambda}^s$ when $\boldsymbol{\lambda} \in int(\mathcal{F}(\mathbf{A}, \mathbf{d}))$, and it is equal to a selection of the supply correspondence when $\boldsymbol{\lambda}$ is in the boundary of $\mathcal{F}(\mathbf{A}, \mathbf{d})$.

Assumption 3 $\mathcal{F}(\mathbf{A}_0, \mathbf{d}) \neq \emptyset$. *That is, the set of feasible labor distributions is non-empty.*

The goal of the Social Planner D is to choose the labor distribution $\boldsymbol{\lambda}$ (by implicitly allowing the immigrant flows $\{I_i\}_{i=1}^m$) to maximize his contemporaneous citizens' well-being. Formally,

Assumption 4 *Every Social Planner D_t has the following Social Welfare function as his objective function:*

$$\sum_{i=1}^{m} \eta_{it} U_{it} \tag{6}$$

and his goal is to maximize it by choosing the labor distribution $\boldsymbol{\lambda}_t$.

The following example shows a concrete case:

Example 2.1 *If future generations do not matter, i.e., $\beta = 0$, then natives' Total Utility at time t, would be $U(\{c_s\}_{s=t}^{\infty}) = u(c_t)$. The wage of a native worker of type i given the worker distribution $\boldsymbol{\lambda}_t^*$ would be $w_{i,t} = F_i(\mathbf{A}_t\boldsymbol{\lambda}_t^*)$, and since $\beta = 0$ we have that optimal consumption $c_t^* = w_{i,t}$, so $u(c_t^*) = u(F_i(\mathbf{A}_t\boldsymbol{\lambda}_t^*))$. Hence, at the beginning of period t, the Social Planner D_t chooses immigration policy $\boldsymbol{\lambda}_t^*$ to maximize the weighted sum of the natives' utilities:*

$$\boldsymbol{\lambda}_t^* = \underset{\boldsymbol{\lambda} \in \mathcal{F}(\mathbf{A}_t, \mathbf{d})}{\operatorname{argmax}} \left\{ \sum_{i=1}^{m} \eta_{it} u(F_i(\mathbf{A}_t \boldsymbol{\lambda}_t)) \right\} \tag{7}$$

where η_{it} is the share of natives with skill i at time t.

Such a myopic economy will tend to benefit each period's most popular group $\eta_{popular,t} = \max_{i \in \mathcal{M}} \{\eta_{it}\}$ and thus very likely exhibit oscillating behavior.

E.g., consider $\mathcal{M} = \{1, 2\}$, $\mathbf{A}_t = \mathbf{I}\ \forall t$, where $\mathbf{I}$ is the 2×2 identity matrix; and let the production function be symmetric, i.e., let $F(\hat{L}_1, \hat{L}_2) = F(\hat{L}_2, \hat{L}_1)\ \forall \hat{L}_1, \hat{L}_2 > 0$. Then starting from $\boldsymbol{\eta}_0 = (1/10, 9/10)$ the resulting equilibrium path will in general satisfy that $\eta_{1,t}^* > 1/2$ whenever t is odd, and $\eta_{1,t}^* > 1/2$ whenever t is even. The higher the risk-aversion (i.e., the higher b_2 as defined in assumption 1), the more this oscillation will dampen. For high enough risk-aversion, the economy will converge to an equal distribution of high and low skill natives, that is, $\lim_{t\to\infty} \boldsymbol{\eta}_t = (1/2, 1/2)$. On the other hand, with a zero risk-aversion, the result is a cyclical path very similar to one of the available equilibria in Ortega's (2005) model.

Now, consider the previous illustration but change the technological process from "$\mathbf{A}_t = \mathbf{I}\ \forall t$" to "$\mathbf{A}_t = diag(1,\ 1.1^t)\ \forall t$," therefore having a skill-biased technological change (i.e., *type $i = 2$ workers experience an increase in their productivity in absolute and relative terms—relative to $i = 1$). Again, this economy will generally exhibit oscillations, and for a high enough risk-aversion it will converge to a degenerate native distribution with all the natives in the upper skill; that is,* $\lim_{t\to\infty} \boldsymbol{\eta}_t = (1, 0)$. ***End of example***

2.2.3 Schooling

Parents may spend resources to affect their offspring's likelihood of acquiring skills (skill acquisition is a random event). For simplicity *we assume that children acquire at least their parent's skill and at most the next higher skill.*[17] By convention, and for the sake of notational simplicity, we let the skill $m + 1$ equal the skill m. Rational behavior implies that worker of type m will (trivially) spend no resources on schooling.

If the child acquires only the parent's skill we call that event an *unsuccessful jump*. If he acquires the next higher skill we call that event a *successful jump*. Since these two are the only possible outcomes it is natural to consider each jump as a *Bernoulli random variable* B, which is independently distributed across families and time (i.e., across (j, t)).

We let B's probability of success be a function of the training expenditures done by the parent and the cost of schooling for the sought-after skill.

Thus a child, whose parent j of type i spends resources s_j on his education, will acquire the skill $i + B(s_j/z_{i+1})$, where $B(s_j/z_{i+1})$ is a Bernoulli random variable with $\Pr[B(s_j/z_{i+1}) = 1] = P(s_j/z_{i+1})$; the function $P(\cdot)$ will be defined below. The type-specific variable z_{i+1} is the training cost for skill $i + 1$, which we assume proportional to the parent's wage: $z_{i+1} = \varsigma_{i+1} w_i$.[18] The function $P(\cdot)$ is defined as follows:

Assumption 5 *P is an increasing, strictly concave, continuously-differentiable function with $P'(0) = \infty$, $P(\infty) = 1$, $P(0) = 0$ and $1 > P \geq 0$.*

This will ensure that all parents spend some resources on their child's schooling whenever the next skill enjoys a higher well-being. At the aggregate/economy-wide level, it is the measure of successful jumps that counts. Therefore we will introduce an aggregate statistic:

[17] A generalized version of this mechanism can be done. Suffice to say that as long as the (median) child does at least as well as his parents, our simplified version will be able to exhibit the same qualitative properties as the generalized case.

[18] This assumption is compatible with many interpretations; here are two: first, it could be that a good amount of the training is done by the parents; second, it could be that this reflects the opportunity cost (in terms of foregone wages) for the "young adult" by training "any further" than his inherited skill

Definition 3 *θ_i is the proportion of attempted jumps, from skill i to skill $i+1$, that are successful. (If there are no jump attempts θ_i is considered zero by convention):*

$$\theta_i = \theta_i\left(\{s_j\}_{j\in\mathcal{L}_i}\right) = \begin{cases} \frac{\mu(\{j\in\mathcal{L}_i |\, B(s_j/z_{i+1})=1\})}{\mu(\{j\in\mathcal{L}_i |\, s_j>0\})} & \textit{for} \quad \mu\left(\{j \in \mathcal{L}_i |\, s_j > 0\}\right) > 0 \\ 0 & \textit{otherwise} \end{cases}$$

where $\mu(\mathcal{S})$ is the measure of the set $\mathcal{S}$. Since $i = 1$ is the lowest skill, the cost z_1 is immaterial. Along similar lines, the success rate θ_m is vacuous since m is the highest skill. Hence the schooling cost vector is $\mathbf{z} \in \mathbb{R}_{++}^{m-1}$ and the success-rate vector is $\boldsymbol{\theta} \in [0,1]^{m-1}$.

2.2.4 Laws of Motion

The technological state A We first define the transition function, or Law of Motion, for technology $\mathbf{A}$, whose process is exogenous.

Assumption 6 *The Law of Motion for Technology follows a (deterministic) Markov process given by $\mathbf{A}_{t+1} = \mathbf{G}\mathbf{A}_t$ where $\mathbf{G} \in \mathbb{R}_{++}^{d(m)}$.*

Definition 4 *The technological process $\{\mathbf{A}_t\}_{t=0}^{\infty}$ is said to be neutral iff $g_i = g$ for all i, for some $g > 0$.*

Definition 5 *The technological process is said to be skill-biased iff $g_i > g_{i'} > 0$ whenever $i > i'$.*

We focus exclusively on these two families of technological processes.

The native distribution $\boldsymbol{\eta}$. Following the structure of our model, there are two parts determining the Motion for the native distribution. One is the immigration policy which transforms the native distribution $\boldsymbol{\eta}_t$ into the worker distribution $\boldsymbol{\lambda}_t$ employed in production. This transformation is chosen by the contemporaneous Social Planner D_t.

The other part determining the future native distribution is the social/educational mobility. Recall that the success rate $\theta_{it} \in [0,1)$ determined the percentage of children, from parents of type i at time t, that became adults of type $i+1$ at period $t+1$. Let $\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t}, ..., \theta_{m-1t})$. Then, the next period's native distribution will be just this period's labor distribution $\boldsymbol{\lambda}_t$ transformed according to $\boldsymbol{\theta}_t$; a linear transformation. We can therefore write the (endogenous) *Law of Motion for $\boldsymbol{\eta}$* as:

$$\boldsymbol{\eta}_{t+1} = \mathbf{M}(\boldsymbol{\theta}_t)\boldsymbol{\lambda}_t \tag{8}$$

where $\mathbf{M}(\boldsymbol{\theta}_t)$ is a $m \times m$ transition matrix, some of whose entries depend on $\boldsymbol{\theta}_t$ and all whose rows add up to one.

Notice $\boldsymbol{\lambda}_t$ depends on the Social Planner D_t and $\boldsymbol{\theta}_t$ depends on the schooling decisions $\{s_j\}_{j\in\mathcal{L}_t}$, but we wrote them here as constants for expositional purposes.

Example 2.2 *Let $m = 3$, then at the beginning of period t_0 there will be a known state, let it be* $(\mathbf{A}_{t_0}, \boldsymbol{\eta}_{t_0}) = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 2/5 \\ 1/5 \\ 2/5 \end{bmatrix} \right)$. *The contemporaneous Social Planner D_{t_0} will choose $\boldsymbol{\lambda}_{t_0}$ to maximize the welfare of these two native groups. Therefore he chooses $\boldsymbol{\lambda}^*_{t_0}$ to maximize the Utility of both groups, weighted by their share of the population. Then, all workers spend an (individually-optimal) amount of resources in their children's education, which translate into a success rate vector $\boldsymbol{\theta}^*_{t_0}$. For the sake of exposition, suppose that* $\boldsymbol{\lambda}^*_{t_0} = \begin{bmatrix} 3/5 \\ 1/5 \\ 1/5 \end{bmatrix}$ *and* $\boldsymbol{\theta}^*_{t_0} = \begin{bmatrix} \theta^*_{1,t_0} \\ \theta^*_{2,t_0} \end{bmatrix} = \begin{bmatrix} 1/4 \\ 2/4 \end{bmatrix}$ *, then the transition matrix is*

$$\mathbf{M}(\boldsymbol{\theta}_{t_0}) = \begin{bmatrix} 1-\theta^*_{1,t_0} & 0 & 0 \\ \theta^*_{1,t_0} & 1-\theta^*_{2,t_0} & 0 \\ 0 & \theta^*_{2,t_0} & 1 \end{bmatrix} = \begin{bmatrix} 3/4 & 0 & 0 \\ 1/4 & 2/4 & 0 \\ 0 & 2/4 & 1 \end{bmatrix}$$

and the native distribution is

$$\begin{aligned} \boldsymbol{\eta}_{t_0+1} &= \mathbf{M}(\boldsymbol{\theta}_{\mathbf{t_0}})\boldsymbol{\lambda}^*_{t_0} \\ &= \begin{bmatrix} 1-\theta^*_{1,t_0} & 0 & 0 \\ \theta^*_{1,t_0} & 1-\theta^*_{2,t_0} & 0 \\ 0 & \theta^*_{2,t_0} & 1 \end{bmatrix} \begin{bmatrix} \lambda^*_{1,t_0} \\ \lambda^*_{2,t_0} \\ \lambda^*_{3,t_0} \end{bmatrix} = \begin{bmatrix} 3/4 & 0 & 0 \\ 1/4 & 2/4 & 0 \\ 0 & 2/4 & 1 \end{bmatrix} \begin{bmatrix} 3/5 \\ 1/5 \\ 1/5 \end{bmatrix} \\ &= \begin{bmatrix} 9/20 \\ 5/20 \\ 6/20 \end{bmatrix} \end{aligned}$$

Notice that there was no need to specify the size of the immigrant flow. The interested reader can verify that, in this particular example (which entails big changes in the skill composition), the immigrant flow would have to be at least 40% of the native population. ***End of example***

Finally we consider the Law of Motion for skill.

Dynasty j's skill i_t Consider dynasty j. Following Subsection 2.2.3 we can write dynasty j's *Law of Motion for skill* as

$$i_{t+1} = i_t + B(s_{jt}/z_{i_t+1}) \quad \forall t$$

this will turn out to be the only truly random Law of Motion, since the aggregate Laws of Motion will, in equilibrium, be deterministic (i.e., there will be no aggregate uncertainty).

2.3 Equilibrium

Let $\bar{\mathbf{\Lambda}}(\mathbf{A}, \boldsymbol{\eta})$ and $\bar{\mathbf{\Theta}}(\mathbf{A}, \boldsymbol{\eta})$ be two given continuous functions with $\bar{\mathbf{\Lambda}} : \mathbb{R}^{d(m)}_{++} \times \Delta^{m-1} \longrightarrow \Delta^{m-1}$ and $\bar{\mathbf{\Theta}} : \mathbb{R}^{d(m)}_{++} \times \Delta^{m-1} \longrightarrow [0, 1)^{m-1}$. These functions are to be considered as the *continuation policies*, i.e., the behavior of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ that will take place in the future or in today's economy and cannot be affected unilaterally at the current time period.

We will make use of the following Laws of Motion (9a-c) and equality (9d), all defined for time $h \geq t$:

$$\mathbf{A}_{h+1} = \mathbf{G}\mathbf{A}_h \quad h = t, t+1, ... \tag{9a}$$

$$\boldsymbol{\eta}_{h+1} = \mathbf{M}(\bar{\mathbf{\Theta}}(\mathbf{A}_h, \boldsymbol{\eta}_h))\boldsymbol{\lambda}_h \quad h = t, t+1, ... \tag{9b}$$

$$i_{h+1} = i_h + B(x_h/z_{i_h+1}) \quad h = t, t+1, ... \tag{9c}$$

$$\boldsymbol{\lambda}_h = \begin{cases} \boldsymbol{\lambda}_t & \text{for} \quad h = t \\ \bar{\mathbf{\Lambda}}(\mathbf{A}_h, \boldsymbol{\eta}_h) & \text{for} \quad h > t \end{cases} \tag{9d}$$

For an individual with skill i_t let $\mathcal{X}_{i_t,t}$ be the set of all *feasible* type and state-contingent plans for schooling expenditures beginning at time t:

$$\mathcal{X}_{i_t,t}(\boldsymbol{\lambda}_t) = \{\{x_{i,h}\}_{i=i_t,h=t}^{m,\infty} \mid x_{i,h} \in [0, F_i(\mathbf{A}_h\boldsymbol{\lambda}_h)] \text{ subject to eqs. } 9a - 9d\}$$

then we define the *planner-dependant Value Function* for such an individual as:

$$\tilde{V}_{i_t}(\mathbf{A}_t, \boldsymbol{\eta}_t, \boldsymbol{\lambda}_t | \, \bar{\mathbf{\Lambda}}, \bar{\mathbf{\Theta}}) = \tag{10}$$

$$\max_{\mathbf{x} \in \mathcal{X}_{i,t}} \left\{ u(F_{i_t}(\mathbf{A}_t\boldsymbol{\lambda}_t) - x_{i_t,t}) + E\left[\sum_{h=t}^{\infty} \beta^{h-t} u(F_{i_h}(\mathbf{A}_h\boldsymbol{\lambda}_h) - x_{i_h,h})\right]\right\}$$

$$s.t. (9a - d)$$

where the expectation operator is over the set of skill sequences $\{i_h\}_{h=t}^{\infty}$ conditional on

$\mathbf{x} \in \mathcal{X}_{i,t}$. Denote the first element $x^*_{it,t}$ of the maximizing plan $\mathbf{x}^* \in \mathcal{X}_{it,t}(\boldsymbol{\lambda}_t)$ for (j,i,t) as $s^*_{jit}(\mathbf{A}_t, \boldsymbol{\eta}_t, \boldsymbol{\lambda}_t | \bar{\mathbf{\Lambda}}, \bar{\mathbf{\Theta}})$, or simply s^*_{jit}, which is j's optimal schooling expenditure at time t.

Thus the planner-dependant Value Function is the maximum Total Utility a worker could get if today's Social Planner enforced immigration policy $\boldsymbol{\lambda}$ and everyone else behaved according to the continuation policies and Laws of Motion ($\bar{\mathbf{\Lambda}}$, $\bar{\mathbf{\Theta}}$ and eqs. 9a-c).

2.3.1 Equilibrium Definition

Definition 6 *Let the model satisfy our assumptions. If:*

1. $\mathbf{\Lambda}(\mathbf{A}_t, \boldsymbol{\eta}_t) = \underset{\boldsymbol{\lambda} \in \mathcal{F}(\mathbf{A}_t, \mathbf{d})}{\arg\max} \left\{ \sum_{i=1}^{m} \eta_{it} \tilde{V}_i(\mathbf{A}_t, \boldsymbol{\eta}_t, \boldsymbol{\lambda} | \, \bar{\mathbf{\Lambda}}, \bar{\mathbf{\Theta}}) \right\} \quad \forall t$

2. $\mathbf{\Theta}(\mathbf{A}_t, \boldsymbol{\eta}_t) = \boldsymbol{\theta}(\{s^*_j\}_{j \in \mathcal{L}_t}) \quad \forall t$

and
$\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$
$\mathbf{\Theta} = \bar{\mathbf{\Theta}}$
then $\langle \mathbf{\Lambda}, \mathbf{\Theta} \rangle$ *is an equilibrium in the Sequential Problem (SP).*

In other words, the immigration policy $\mathbf{\Lambda}$ and the success rate function $\mathbf{\Theta}$ are an equilibrium if the former is what a rational Social Planner would choose, the latter is the actual success rate given workers optimal schooling choices, and if these two are equal to the behavior anticipated by everyone.

2.3.2 Existence and Uniqueness

Let $\mathrm{int}(\mathcal{S})$ denote the interior of set $\mathcal{S}$. Then we have:

Theorem 1 (Existence and Uniqueness) *Let the model satisfy our assumptions, then there is a value* $\bar{\beta} \geq 0$ *such that if* $0 \leq \beta \leq \bar{\beta}$, *an equilibrium in the Sequential Problem (SP) exists. Moreover, if the composite functions* $h_i \equiv u \circ F_i \; \forall i$ *are strictly concave, then the equilibrium is unique.*

Proof: In appendix B.2

It is worth explaining the relationship between the strict concavity of the functions $h_i(\cdot)$ to the instant utility function $u(\cdot)$ and the production function $F(\cdot)$. One can interpret this condition loosely as follows: workers prefer the certainty equivalent to a "lottery" that gives them a high and a low wage. Of course the condition specifies the details of *how high* and *how low* this wages of the lottery are based on the properties of the production function; and it also specifies the details of exactly how risk averse the workers have to be to satisfy the condition (some risk aversion is clearly necessary though).

2.3.3 Recursive Representation

Theorem 2 (Recursive Representation) *The equilibrium $\langle \mathbf{\Lambda}, \mathbf{\Theta} \rangle$ satisfies that for all i, the functions $V_i(\mathbf{A}, \boldsymbol{\eta}) \equiv \tilde{V}_i(\mathbf{A}, \boldsymbol{\eta}, \mathbf{\Lambda}(\mathbf{A}, \boldsymbol{\eta})|\ \mathbf{\Lambda}, \mathbf{\Theta})$, $\mathbf{\Lambda}$ and $\mathbf{\Theta}$ are continuous and are given by eqs. 11-13:*

$$\begin{gathered} V_i(\mathbf{A}, \boldsymbol{\eta}) = \\ \max_{s \in [0, F_i(\mathbf{A}\mathbf{\Lambda}(\mathbf{A},\boldsymbol{\eta}))]} \left\{ \begin{array}{c} u\left(F_i(\mathbf{A}\mathbf{\Lambda}(\mathbf{A}, \boldsymbol{\eta})) - s\right) \\ +\beta E_B\left[V_{i+B}(\mathbf{A}', \boldsymbol{\eta}') \mid B = B(s/z_{i+1})\right] \end{array} \right\} \\ s.t. \quad \mathbf{A}' = \mathbf{A}\mathbf{G} \\ and\ \boldsymbol{\eta}' = \mathbf{M}(\mathbf{\Theta}(\mathbf{A}, \boldsymbol{\eta}))\mathbf{\Lambda}(\mathbf{A}, \boldsymbol{\eta}) \end{gathered} \tag{11}$$

$$\begin{gathered} \mathbf{\Lambda}(\mathbf{A}, \boldsymbol{\eta}) = \\ \arg\max_{\boldsymbol{\lambda} \in \mathcal{F}(\mathbf{A}, \mathbf{d})} \left\{ \sum_{i=1}^{m} \eta_i \left(\begin{array}{c} u(F_i(\mathbf{A}\boldsymbol{\lambda}) - z_{i+1} P^{-1}(\Theta_i(\mathbf{A}, \boldsymbol{\eta}))) \\ +\beta E_B\left[V_{i+B}(\mathbf{A}', \boldsymbol{\eta}') \mid B = B\left(P^{-1}(\Theta_i(\mathbf{A}, \boldsymbol{\eta}))\right)\right] \end{array} \right) \right\} \\ s.t. \quad \mathbf{A}' = \mathbf{A}\mathbf{G} \\ and\ \boldsymbol{\eta}' = \mathbf{M}(\mathbf{\Theta}(\mathbf{A}, \boldsymbol{\eta}))\mathbf{\Lambda}(\mathbf{A}, \boldsymbol{\eta}) \end{gathered} \tag{12}$$

$$\begin{gathered} \Theta_i(\mathbf{A}, \boldsymbol{\eta}) = P(s_i^*(\mathbf{A}, \boldsymbol{\eta}))\ \forall i \\ \textit{where } s_i^*(\mathbf{A}, \boldsymbol{\eta}) \textit{ is the maximizer of eq. 11} \end{gathered} \tag{13}$$

together called the Functional Equation System (FES).

Proof: In Appendix B.3

2.4 Immigration policies

Assume the uniqueness conditions are satisfied for the rest of the paper.

Given that the equilibrium is unique and satisfies the recursive representation we are able to characterize the equilibrium immigration policy $\mathbf{\Lambda}$. It is especially useful to compare $\boldsymbol{\lambda}$, the worker distribution (in equilibrium equal to $\mathbf{\Lambda}$), with $\boldsymbol{\eta}$, the native distribution, thus revealing in which direction and at what skill levels the immigrant skill composition diverges from the native skill composition. A stronger version of this statistic is the difference between immigrant $\boldsymbol{\iota}$ and native distributions $\boldsymbol{\eta}$.[19] It is easy to show that the three distributions have the following relationship:

$$\boldsymbol{\lambda} = a\,\boldsymbol{\eta} + (1-a)\boldsymbol{\iota}$$

where $a = N/L$. Clearly, if $\boldsymbol{\lambda}$ differs from $\boldsymbol{\eta}$ then it does so in exactly the same direction as $\boldsymbol{\iota}$, but since $\boldsymbol{\lambda}$ is a convex combination of both distributions, it is generally a weaker statistic than $\boldsymbol{\iota}$. Hence our use, especially in the empirical Section of the paper, of the following statistic:

Definition 7 *$\boldsymbol{\phi} \equiv \boldsymbol{\iota} - \boldsymbol{\eta}$ is the difference between immigrant and native skill distributions.*

Phi ($\boldsymbol{\phi}$) provides information on which, and by how much, skills are "boosted" or "suppressed" with respect to the native distribution. Subsection 3.1 shows smoothed versions of Phi for the USA and Canada for the available Censuses. For example, $\phi_m > 0$ means $\iota_m > \eta_m$—immigrants have a higher percentage of top skill workers than natives.

Our main results are in the following two Main Theorems.

2.4.1 Neutral Technological Change

Theorem 3 (Main Theorem on Skill-Neutral Technological Change) *Let there be a unique equilibrium under neutral technological change and with positive discount factor, that is, with $\mathbf{G} \equiv g \underset{m\times m}{\mathbf{I}}$ for some $g > 0$, and with $\beta > 0$. Then, there exists a stationary native distribution, call it $\boldsymbol{\eta}^*_{SS}$, which has an associated stationary worker distribution $\boldsymbol{\lambda}^*_{SS}$, and an associated stationary success rate vector $\boldsymbol{\theta}^*_{SS}$. Moreover:*
1) $\boldsymbol{\eta}^*_{SS}$ first order dominates $\boldsymbol{\lambda}^*_{SS}$, i.e., $\phi_{i,SS}$ is decreasing in i.
2) $V_i(\mathbf{A}, \boldsymbol{\eta}^*_{SS}) > V_{i'}(\mathbf{A}, \boldsymbol{\eta}^*_{SS})\ \forall i, i' \in \mathcal{M}$ with $i > i'$, and $\forall \mathbf{A} \in \{\mathbf{A}_t\}_{t=0}^{\infty}$
3) $\boldsymbol{\theta}^*_{SS} > \mathbf{0} \equiv (0, 0, \ldots, 0)$.
4) $F_i(\mathbf{A}\boldsymbol{\lambda}^*_{SS}) > F_{i'}(\mathbf{A}\boldsymbol{\lambda}^*_{SS})\ \forall i, i' \in \mathcal{M}$ with $i > i'$, and $\forall \mathbf{A} \in \{\mathbf{A}_t\}_{t=0}^{\infty}$.

Proof: In Appendix B.4

Theorem 3 says that under neutral technological change, the steady-state native skill distribution will always first-order stochastically dominate the steady-state worker skill distribution and, hence, the steady-state immigrant skill distribution. Thus there will be low-skill immigrants as long as immigration is positive. Moreover, regardless of the

[19] $\boldsymbol{\iota}$ was defined together with $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ on page 10.

metric/scale used for skills, the average skill of immigrants will always be lower than that of natives.

The intuition for this result is as follows: Under isolation (without any possibility of immigration), the economy would tend to perfect equality, i.e., every skill would earn the same wage in the limit (this being the steady state). However the possibility of immigration, and the fact that schooling is decentralized, always encourages some immigration and schooling.

Clearly the high-skill natives are better off, earning an above-the-egalitarian wage and benefiting from their scarcity vis-a-vis the abundance of low skilled workers. As for the rest of the natives, they will become high-skilled in finite time, with probability one, therefore enjoying an above-the-egalitarian wage (being the scarce factor) thereafter.

2.4.2 Skill-biased Technological Change

Theorem 4 (Main Theorem on Skill-Biased Technological Change) *Let there be a unique equilibrium under skill-biased technological change with $m \geq 3$, allowing all—except the bottom—skill outside options d_i with $i > 1$ to increase at a positive (though small), order preserving, rate, i.e., let $g_m > \ldots g_2 > g_{d_m} > \ldots > g_{d_2} > 1 \geq g_{d_1} \lesseqqgtr g_1$. Let $\{\boldsymbol{\eta}_t^*, \boldsymbol{\lambda}_t^*, \boldsymbol{\theta}_t^*\}_{t=0}^{\infty}$ be the equilibrium-path values for $\boldsymbol{\eta}, \boldsymbol{\lambda}$ and $\boldsymbol{\theta}$, then:*
1) There exists no stationary native distribution $\boldsymbol{\eta}_{SS}$.
2) $\boldsymbol{\theta}_t^* > \mathbf{0}$ $\forall t$
3) $V_i(\mathbf{A}_t, \boldsymbol{\eta}_t^*) > V_{i'}(\mathbf{A}_t, \boldsymbol{\eta}_t^*)$ $\forall i, i' \in \mathcal{M}$ with $i > i'$, and $\forall t \in \mathbb{N}$
4) $\phi_{1,t}^* > 0$ for all $t > n_{SB1}$ for some $n_{SB1} \in \mathbb{N}$.
Moreover if the relative risk aversion is strong enough, and the discount factor is small enough, then:
5) $\phi_{m,t}^* > 0$ for all $t > n_{SB2}$ for some $n_{SB2} \in \mathbb{N}$

Proof: In Appendix B.5

Corollary (3 Skills)
If the conditions of Theorem 4 are satisfied and $m = 3$, then $\phi_{i,t}$ is u-shaped in i for all $t > n_C$ for some $n_C \in \mathbb{N}$.

The previous corollary says that in an economy with exactly three skills, skill-biased technological change and strongly risk-averse citizens, the immigration policy will favor the overrepresentation of immigrants at both ends of the skill spectrum, leading to a 3-segment pattern as the one in figure 1 on page 3.

3 Empirical Findings

3.1 Stylized facts

Let us proceed directly to analyze the skill distributions for the USA and Canada through the following figures. These figures are smoothed versions of Phi (the difference between native and immigrant distributions, definition 7 on page 19) for each available Census. A detailed description of their construction follows:

1. We take all the individuals who are 25 through 45 years old from each Census.

2. We classify those who were born outside of the country in question as immigrants, the rest as natives.

3. We get the distribution of educational level for both of these groups and construct the difference between these two (immigrant minus native). This difference is Phi.

4. We smooth Phi using a Hodrick-Prescott filter with smoothness parameter equal to two.

This is done for each and every available Census for each country. We have included the source tables, i.e., the distributions of immigrant and native educational attainment in Appendix C.[20] It is a shame that the USA Censuses did not consider the whole skill spectrum until 1990 (capping it at the college level), since this has obviously distorted our USA figures for the previous time periods (very likely hiding the real pattern, accurately displayed for the years 1990 and 2000). For the sake of completeness, these have been included nevertheless. The resulting figures follow.

[20] Some of the category names were too long to be displayed in the figures. Please see the source tables for the complete category names.

3.1.1 USA

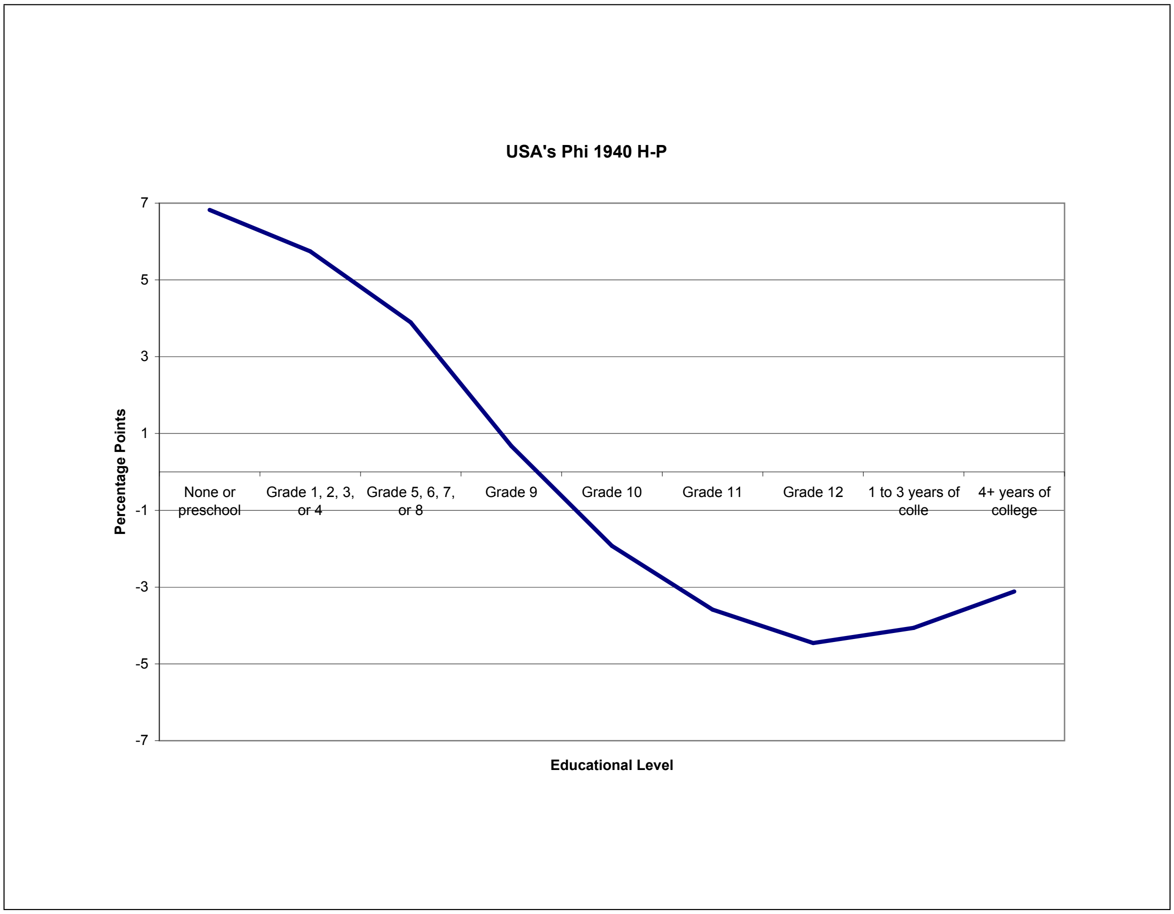


Figure 3: Phi, USA 1940

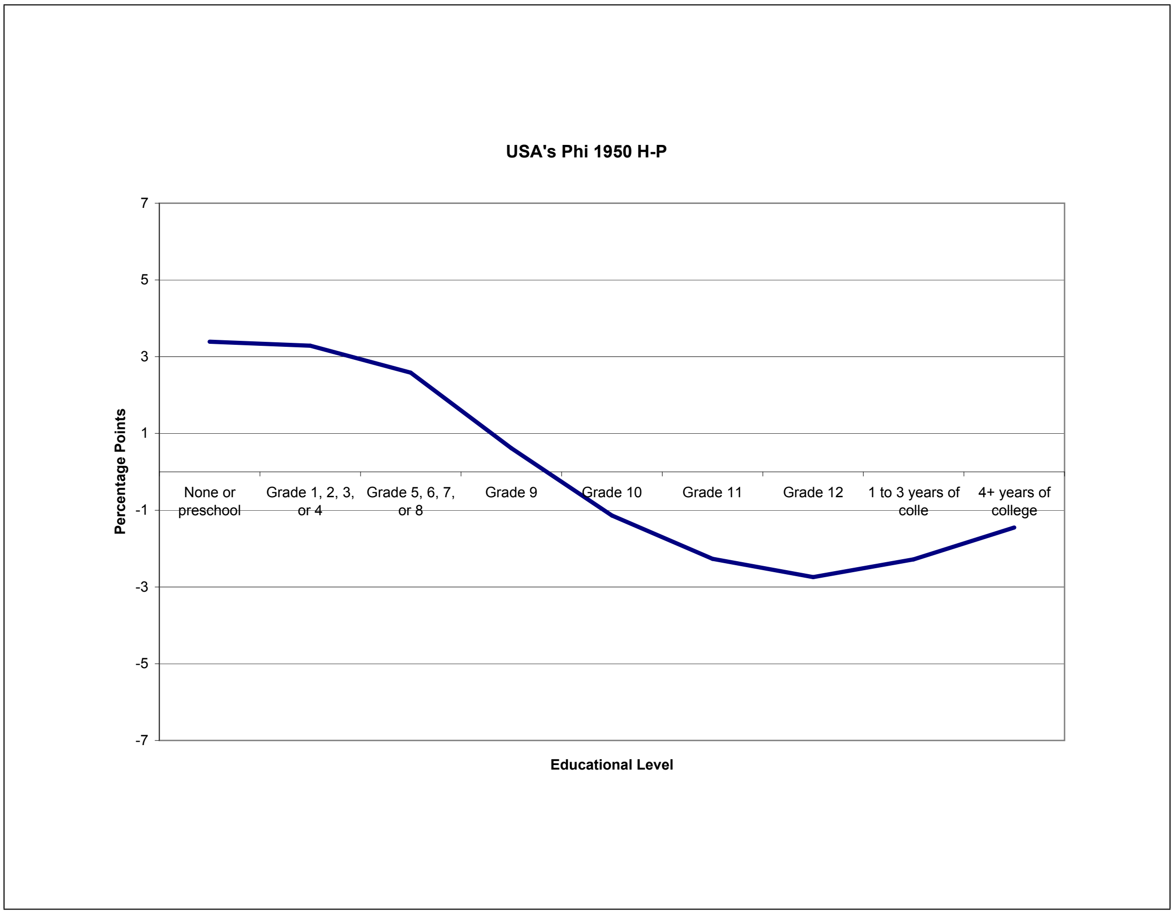


Figure 4: Phi, USA 1950

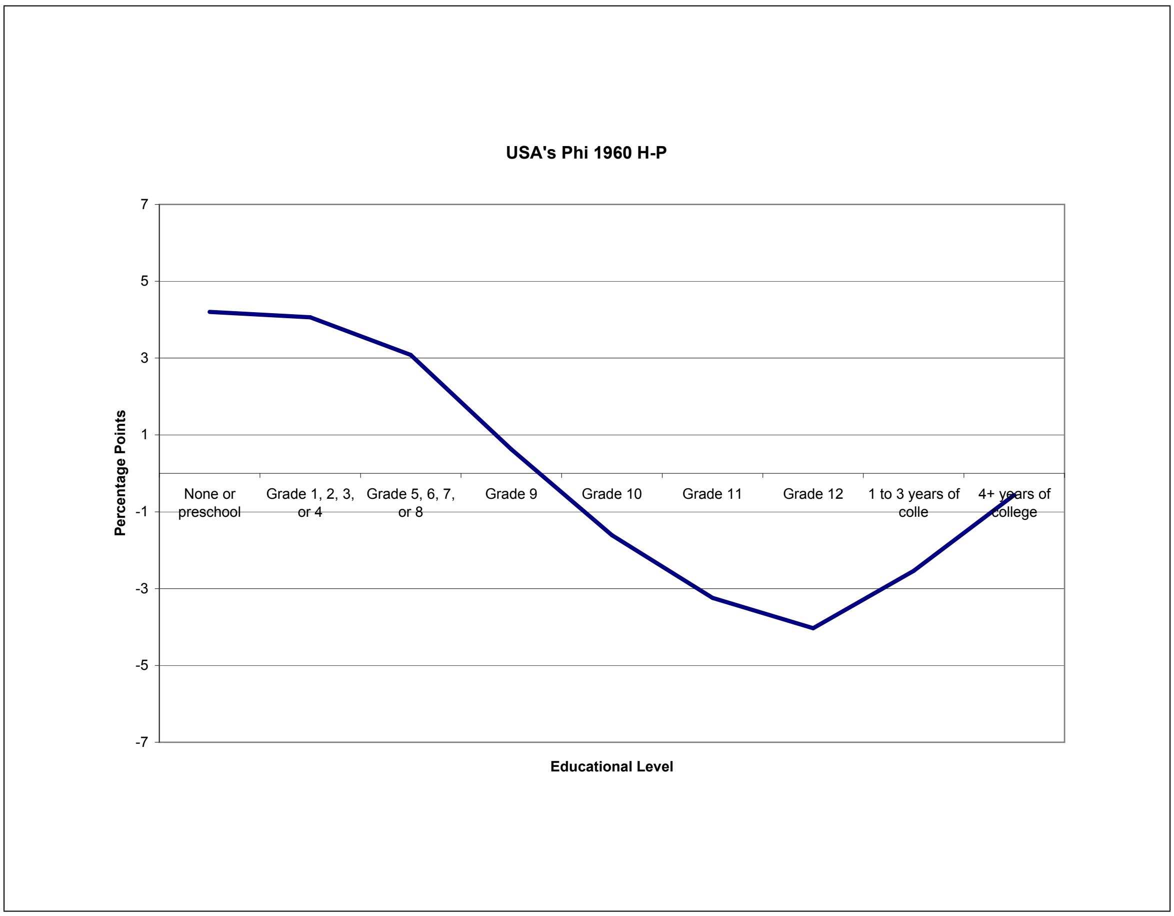


Figure 5: Phi, USA 1960

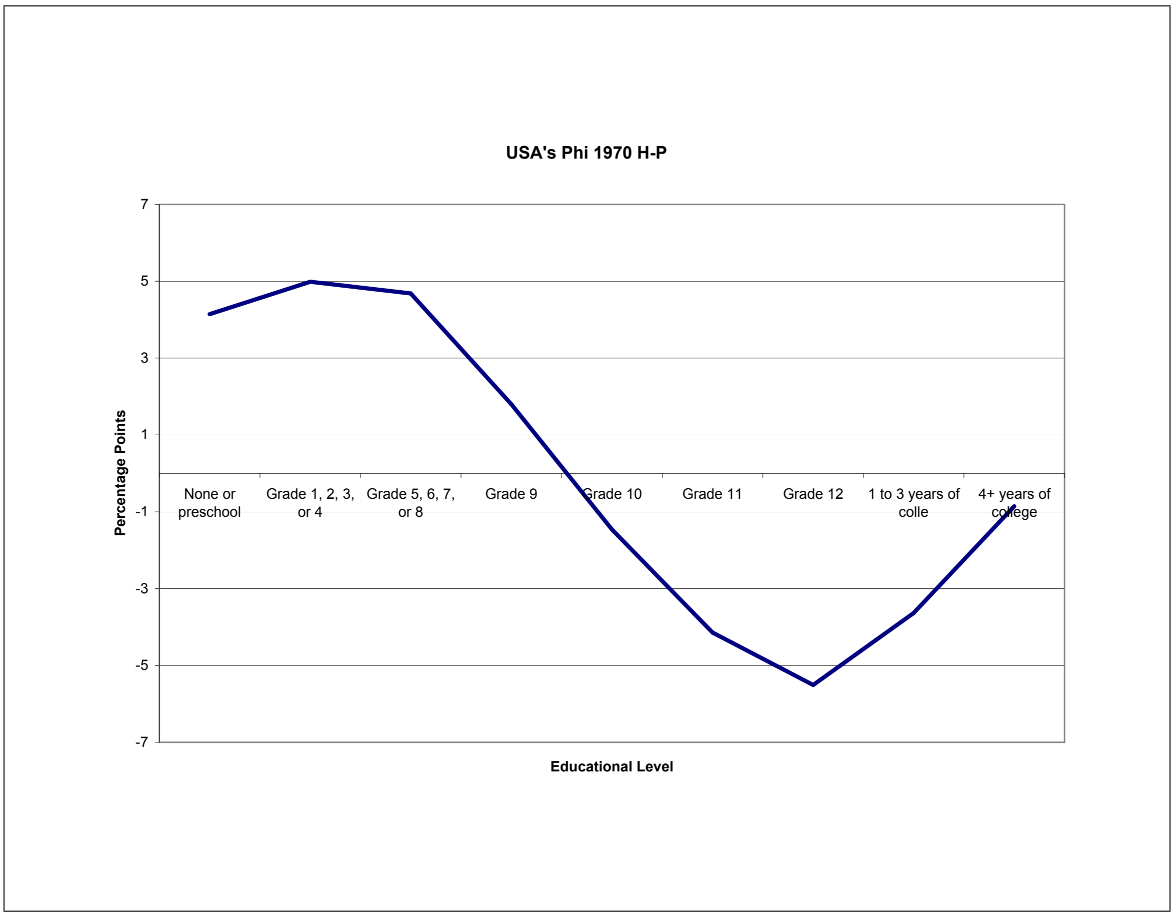


Figure 6: Phi, USA 1970

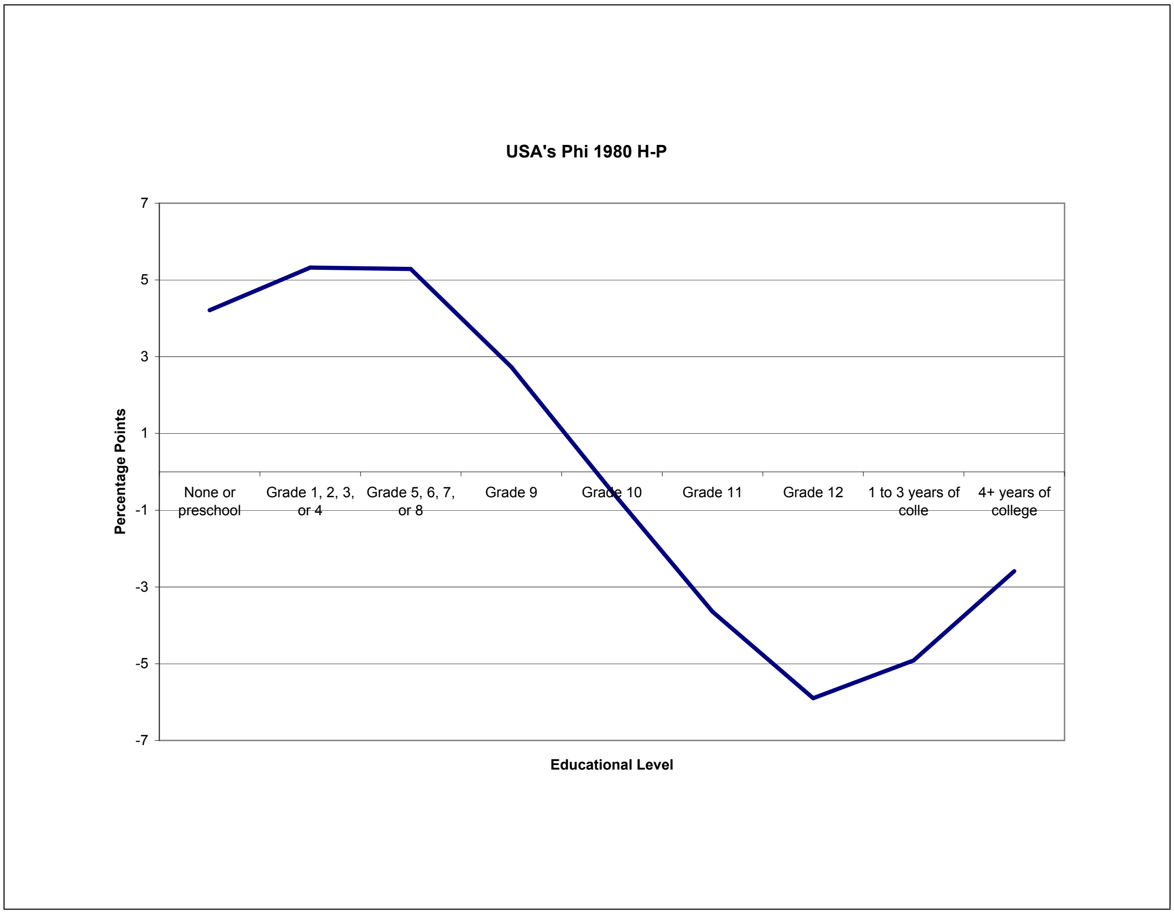


Figure 7: Phi, USA 1980

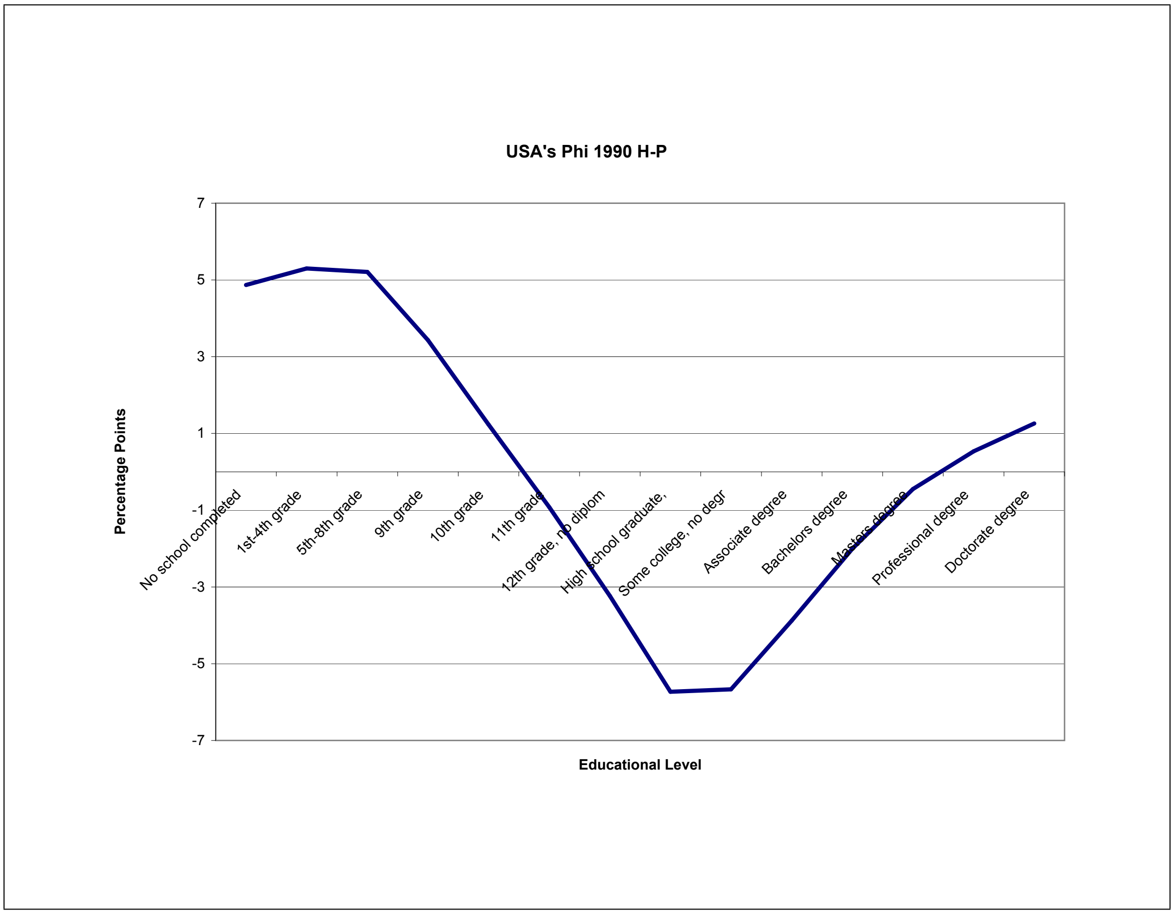


Figure 8: Phi, USA 1990

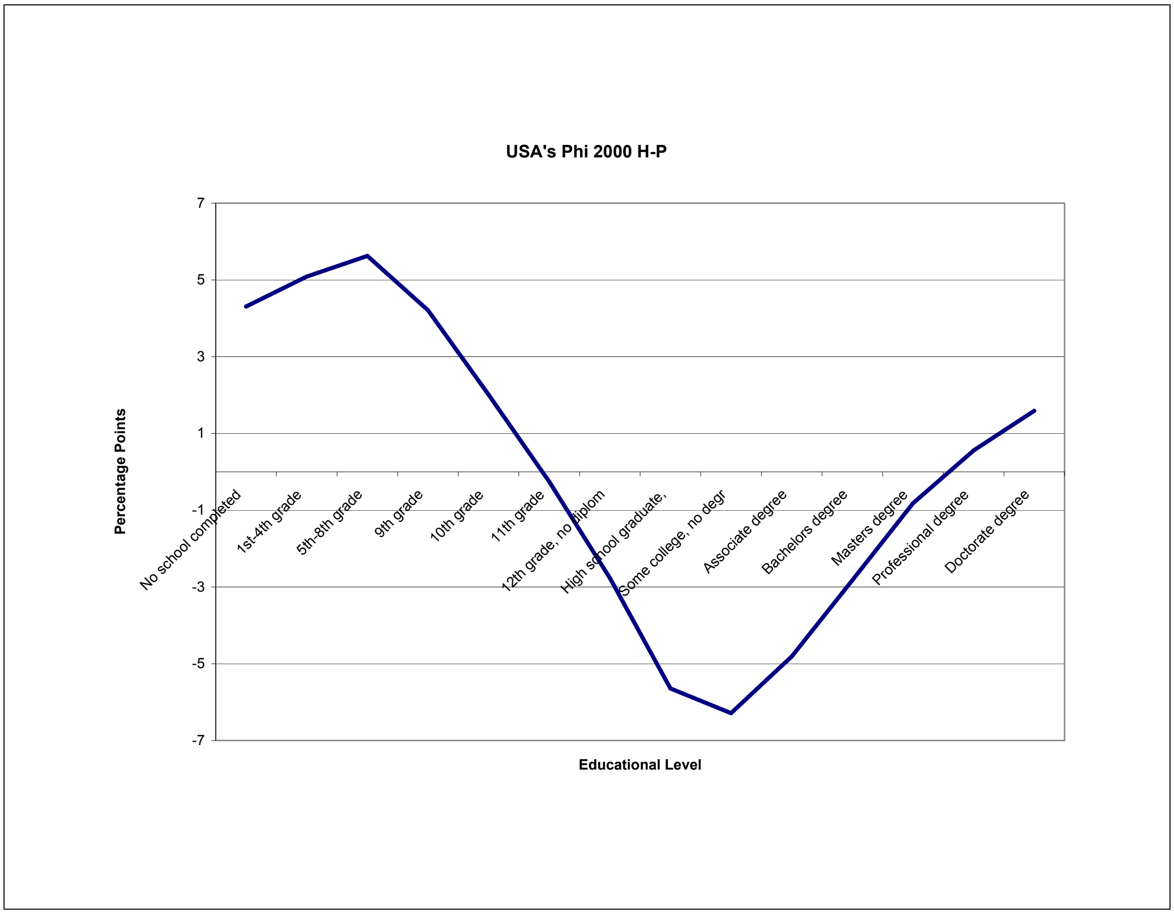


Figure 9: Phi, USA 2000

3.1.2 Canada

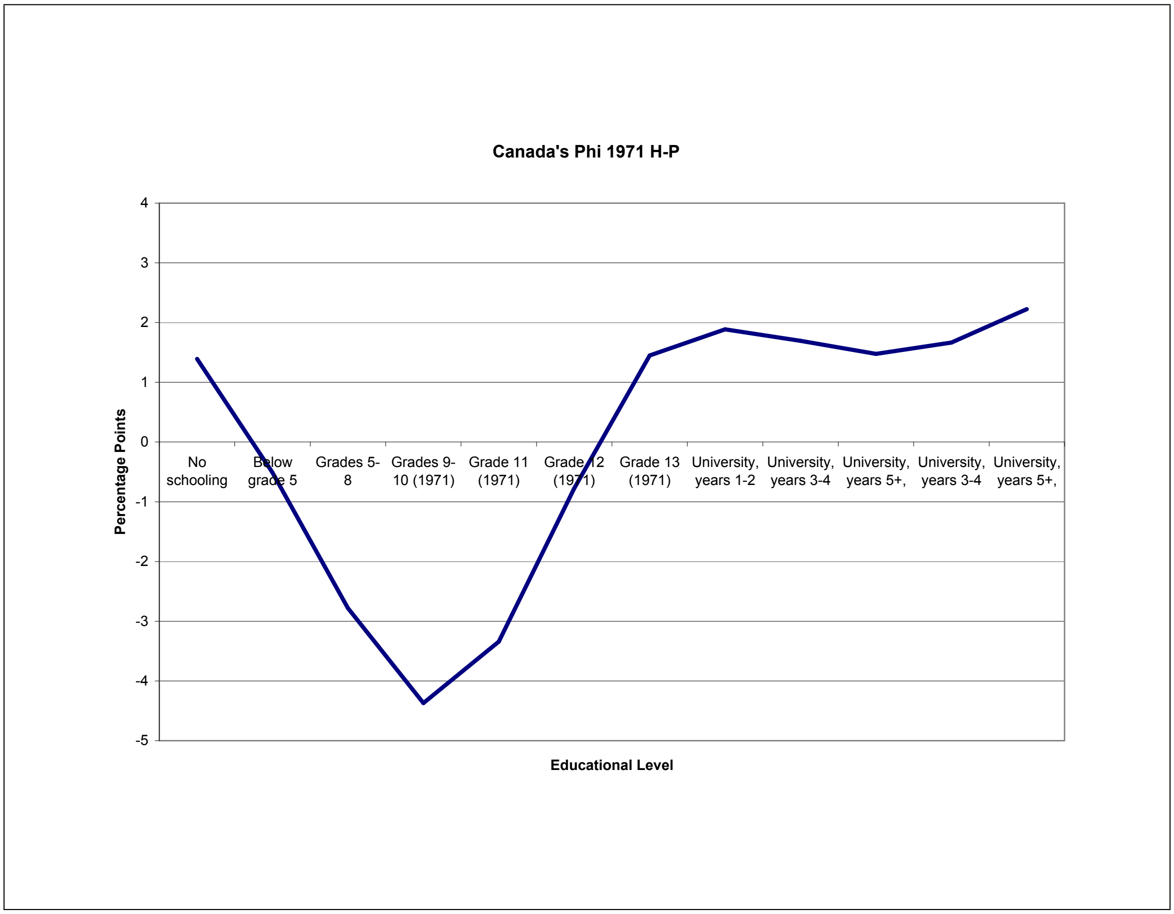


Figure 10: Phi, Canada 1971

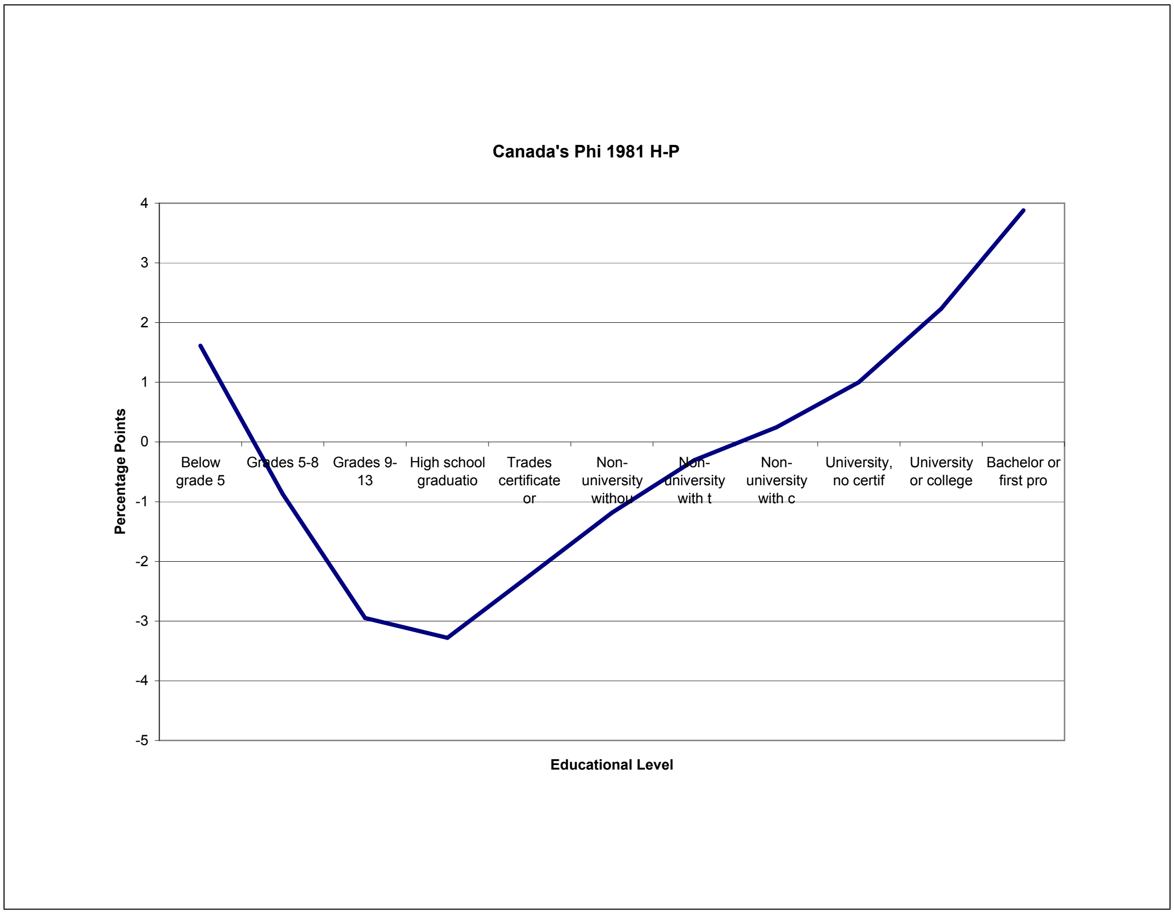


Figure 11: Phi, Canada 1981

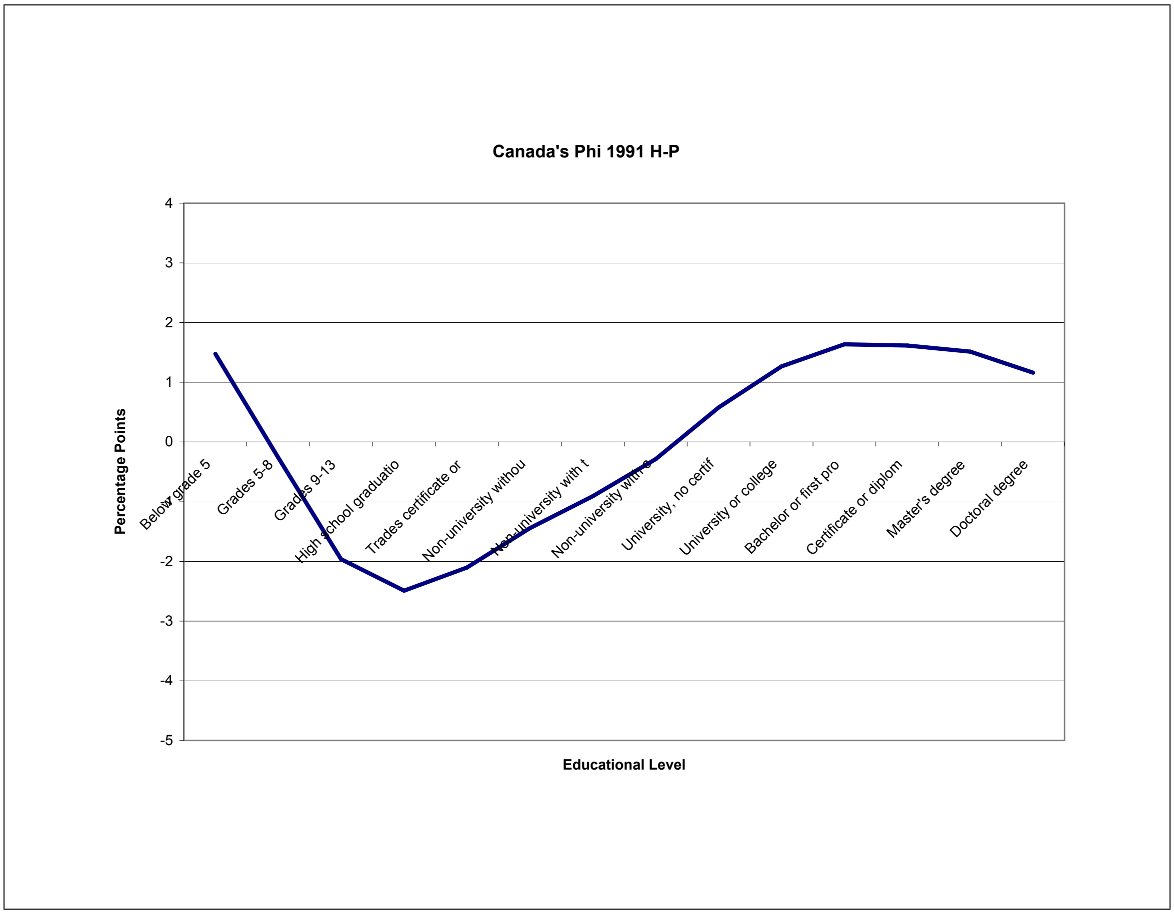


Figure 12: Phi, Canada 1991

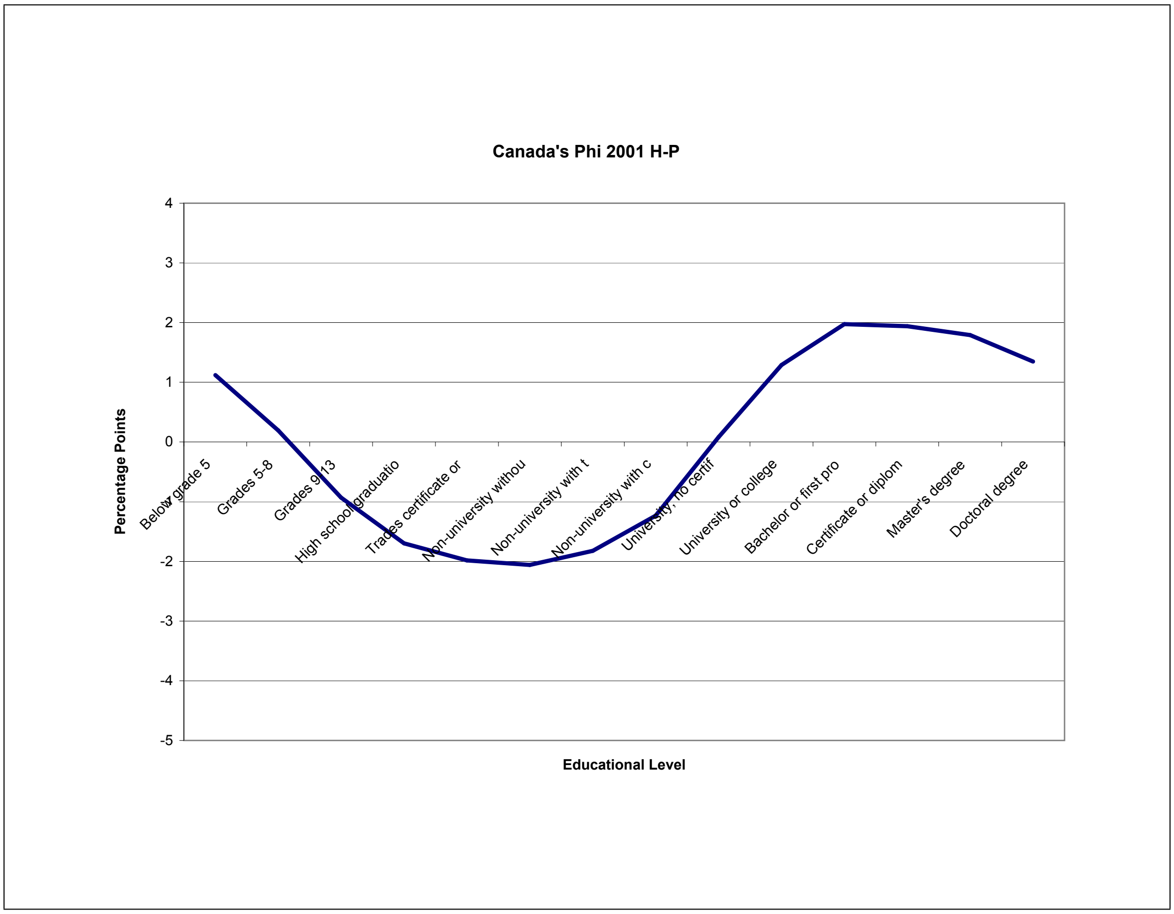


Figure 13: Phi, Canada 2001

3.2 Estimation of the technological process

Observation of the *empirical* Phi in the previous section suggests that the theory will be consistent with the data *only if* both Canada and the USA behave *as if* they experienced skill-biased technological change and had a finite elasticity of substitution greater than one between their different types of labor.

In order to test these two conditions, we proceed to estimate the production function and its technological process. We assume that both the USA and Canada have the same technological process and production function, and estimate them using USA data.

Since the data is generally scarce we need to make some parametric assumptions on F. We therefore use the familiar CES functions.

Another parametric choice that we have to make is the number of skills/schooling levels considered. The trade-offs are, among others, the degrees of freedom versus the accuracy of the technological-process estimation. The minimum number of skills needed to check for consistency of the theory is $m = 3$. Since our available data is a short panel we impose these 3 skills, aggregating as suggested by the 3-segment partition of the schooling spectrum in figure 1.

Our universe for this estimation contains all the individuals who received positive wages for the last year before the Census, and whose age is at least 30 years. We use data from IPUMS USA,[21] which allows us to exploit the disaggregated micro data on schooling attendance for the Censuses 1940 through 2000.

We assume the following specification

$$
\begin{aligned}
F(\hat{\lambda}_{ts}) &= \left[\sum_{i=1}^{3} \left(\hat{\lambda}_{its} \right)^{\rho} \right]^{1/\rho} \\
\hat{\lambda}_{its} &= a_{i,t} \lambda_{its} \\
a_{i,t} &= a_{i,0} \left(g_i \right)^{t}
\end{aligned}
$$

where s stands for political state of the USA (i.e., the cross-sectional unit).

Table 1 displays the estimated terms from two different approaches (with different sets of assumptions).[22] According to Table 2, which displays the implied parameters of interest, both of the elasticity estimates lie within a range of $(1, 8)$, a range that other estimations, like the ones in the survey by Hamermesh (1993), would suggest as standard. *We conclude that the first condition for consistency with our model is satisfied* (finite elasticity greater than one).

[21] Ruggles, S. et al. (2004).

[22] For the sake of space—and since this part is not the main contribution of the paper—the details of the estimation process have been omitted. They are available upon request.

Table 1: Estimates from Specification

Term	Method I	Method II
$\rho\ln(a_{1,1940})$	3.772 318****	NA
	(0.087 496)	
$\rho\ln(a_{2,1940})$	3.825 921****	NA
	(0.120 037)	
$\rho\ln(a_{3,1950})$	3.783 906****	NA
	(0.165 038)	
$\ln(g_1)$	-0.030 374**	-0.076 783****
	(0.014 973)	(0.007 954)
$\ln(g_2)$	0.067 080****	0.023 069****
	(0.003 282)	(0.002 723)
$\ln(g_3)$	0.136 272****	0.027 345****
	(0.006 301)	(0.006 412)
$\rho - 1$	-0.147 203****	-0.941 169****
	(0.021 552)	(0.009 017)

We employ the following markers for the respective levels of confidence: (*)90%, (**)95%, (***)99%, and (****)99.9%. Standard errors are in parentheses.

Table 2: Implied Point Estimates of Interest

Parameter	Method I	Method II
$a_{1,1940}$	83.384 874	NA
$a_{2,1940}$	88.794 301	NA
$a_{3,1950}$	84.525 660	NA
g_1	0.970 082	0.926 090
g_2	1.069 381	1.023 337
g_3	1.145 994	1.027 723
ε	6.793 356	1.062 507

Table 2 displays as well the estimated technological processes. In both cases skill-biased with the point estimates for the bottom-, medium- and top-efficiency growth rates in the ranges [-8%,-2%],[2%,7%] and [3%,14%] respectively. *We conclude that the second condition for consistency with our model is satisfied as well*, namely, these economies behave *as if* they experienced a skilled bias technological change.

In other words, our model is consistent with the data: if the elasticity (of substitution) is greater than one, and the economy behaves as if it experienced skill-biased technological change, then we should observe the patterns from figures 8 through 13.

4 Discussion and Concluding Remarks

This paper has shown evidence of unconventional immigration patterns in two of the traditional countries of destination (USA and Canada) for recent decades. These unconventional immigration patterns consist of overrepresentation of immigrants at both ends of the skill distribution (and hence an underrepresentation of immigrants at the middle of the skill distribution).

The paper also proposed a theoretical framework, capable of generating such patterns under plausible assumptions such as strong risk aversion, credit restrictions, and filial altruism, among others.

In general, the theory predicts that under any skill-biased technological change, the high skilled and low skilled individuals of less developed countries will be attracted by rapidly growing democracies. This in turn would make the skill composition of the source countries more homogeneous and reduce the low-middle skill wage-differential until it is equalized between all, the countries of origin and the countries of destination.

Analysis of other traditional destination countries, such as Australia, as well as of other fast-growing democracies, like New Zealand, remains to be done. When doing such analyses the most-possible-disaggregated data is to be employed, since otherwise the pattern that has been pointed out could be eliminated through aggregation in the same manner it probably was eliminated at the top of the 80s skill distribution for the USA, due to the coarse classification of that Census.

Some of the possible directions for further research include analyzing the effects of given income-tax funded provision of public goods and/or simply the effects of given redistributive schemes. Intuition suggests that such schemes would likely result in less desirability of (higher opposition to) unskilled immigrants.

Finally, the search for another explanation of the exposed phenomenon is open; special-interest politics being a natural candidate. In the author's opinion this explanation is complementary and hence "non-rival." He also strongly believes that any other explanation would need to acknowledge the fact that immigration policies are binding (i.e., entry into fast-growing democracies is—at least partially—controlled) in order to be seriously considered.

A Appendix: Quick Reference for Notation

A.1 Set Theoretical Notation

$\mathcal{S} = \{x, y, z\}$ means $\mathcal{S}$ is a *set* whose elements are x, y, z

$\mathcal{S} = \{x : p(x)\}$ means $\mathcal{S}$ is the set of x's for which statement p (which depends on x) is true; often "|" is used instead of ":"

$x \in \mathcal{S}$ means x is an *element* of the set $\mathcal{S}$

$\mathcal{Q} \subseteq \mathcal{S}$ means $\mathcal{Q}$ is a *subset* of $\mathcal{S}$, i.e., $\mathcal{Q}$ is a set whose elements are all contained in $\mathcal{S}$ as well

$\mathcal{Q} \subset \mathcal{S}$ means $\mathcal{Q}$ is a *proper subset* of $\mathcal{S}$, i.e., $\mathcal{Q}$ is a set whose elements are all contained in $\mathcal{S}$ but who is not equal to $\mathcal{S}$. Hence $\mathcal{Q}$ has some but not all the elements contained in $\mathcal{S}$.

$\mathbb{N}$ is the set of natural numbers: $\mathbb{N} = \{1, 2, 3, ...\}$

$\mathbb{N}_0$ is the set of non-negative integers: $\mathbb{N}_0 = \{0, 1, 2, ...\}$

$\mathbb{R}$ is the set of real numbers: $\mathbb{R} = \{x : +\infty > x > -\infty\}$

$\mathbb{R}_+$ is the set of non-negative real numbers: $\mathbb{R} = \{x : +\infty > x \geq 0\}$

$\mathbb{R}_{++}$ is the set of positive real numbers: $\mathbb{R} = \{x : +\infty > x > 0\}$

$\mathbb{R}^b$ is the set of b-dimensional real vectors, where $b \in \mathbb{N}$, thus: $\mathbb{R}^b = \{\mathbf{x} = (x_1, x_2, ...x_b) : x_k \in \mathbb{R} \text{ for all } k = 1, 2, ...b\}$

$\mathbb{R}^{b \times c}$ is the set of real valued matrices of dimension $b \times c$, where both $b, c \in \mathbb{N}$ thus: $\mathbb{R}^{b \times c} = \{[b_{ij}]_{i=1,j=1}^{b,c} : x_k \in \mathbb{R} \text{ for all } k = 1, 2, ...b\}$

A.2 Notation of the paper

j is the index for families / workers / dynasties, we use $j \in \mathcal{L} \subseteq \mathbb{R}$

i is the index for skill / type, we use $i \in \mathcal{M}$

t is the index for time, we use $t \in \mathbb{N}_0$

m is the total number of skills *and* the index of the top skill

$\mathcal{M}$ is the set of skills, we assume $\mathcal{M} = \{1, 2, ..., m\} \subset \mathbb{N}$

T is the last period in a finite-horizon Setup, that is $T \in \mathbb{N}_0$

$u(c)$ is the *Instant Utility* function, where $c \in \mathbb{R}_+$ is consumption

U_{it} is the *Total Utility* of some i at time t

$F(\mathbf{x})$ is the time-invariant production function, where $\mathbf{x} \in \mathbb{R}^m_+$

$\mathcal{L}$ is the set of workers, we let $\mathcal{L} \subseteq \mathbb{R}$

L is the measure of the set $\mathcal{L}$, formally $L = \mu(\mathcal{L})$, where $\mu(\mathcal{S})$ is called the *Lebesgue measure* of set $\mathcal{S}$.

$\mathcal{L}_i$ is the set of workers with skill i, i.e., $\mathcal{L}_i = \{j \mid j \text{ is of skill } i\}$

L_i $= \mu(\mathcal{L}_i)$

$\mathcal{L}_t$ is the set of workers at time t

L_t $= \mu(\mathcal{L}_t)$

$\mathcal{L}_{it}$ is the set of workers with skill i at time t

L_{it} $= \mu(\mathcal{L}_{it})$

$\mathcal{N}$ is the set of natives, and if immigration is non-negative then $\mathcal{N} \subseteq \mathcal{L}$

N $= \mu(\mathcal{N})$

$\mathcal{I}$ is the set of immigrants

I $= \mu(\mathcal{I})$

$\boldsymbol{\lambda}$ $= (L_1, L_2, ..., L_m)L^{-1}$ is called the *labor distribution*

$\boldsymbol{\eta}$ $= (N_1, N_2, ..., N_m)N^{-1}$ is called the *native distribution*

$\boldsymbol{\iota}$ $= (I_1, I_2, ..., I_m)I^{-1}$ is called the *immigrant distribution*

s is the schooling expenditure variable

z_i is the cost of schooling for skill i

$P(\cdot)$ is the "P-function" a function that gives the probability of acquiring the next higher skill given expenditures over the cost of the sought-after skill.

$B(\cdot)$ is the Bernoulli random variable taking value of one for a successful jump to the next skill and zero if not a successful jump, i.e., the child remains with his parent's skill

$\boldsymbol{\theta}$ is the success rate vector, of dimension $(m-1) \times 1$

$\boldsymbol{\Theta}(\cdot, \cdot)$ is the vector-valued (aggregate-)schooling policy function

$\boldsymbol{\Lambda}(\cdot, \cdot)$ is the vector-valued immigration policy

$\tilde{V}_{it}$ is the *planner-dependant Value Function*, i.e., the Value Function of an individual i at time t as a function of State and policy choice of the Social Planner at time t

V_{it} is the *equilibrium Value Function*

B Appendix: Proofs of Section 2

B.1 Proof of Proposition 1

We want to show that

$$\mathcal{R}_2(\mathbf{A}) \equiv \{\mathbf{x} \in \mathbb{R}^m_+ | \, F_i(\mathbf{A}\mathbf{x}) \geq F_{i-1}(\mathbf{A}\mathbf{x}) \quad \forall i > 1\} \cap \Delta^{m-1}$$

is a convex set.

Definition 8 *The Set of Wage Equality Distributions is*

$$\mathcal{E}(\mathbf{A}) = \{\boldsymbol{\lambda} \in \Delta^{m-1} | F_i(\mathbf{A}\boldsymbol{\lambda}) = F_{i'}(\mathbf{A}\boldsymbol{\lambda}) \forall i, i' \in \mathcal{M}\}$$

which is a non-empty set by Assumption 3.

Definition 9 *the Best Wage Equality Distribution is* $\boldsymbol{\lambda}^{=}_{\mathbf{A}} = \underset{\boldsymbol{\lambda} \in \mathcal{E}(\mathbf{A})}{\arg\max} \; F(\mathbf{A}\boldsymbol{\lambda})$

We just write $\boldsymbol{\lambda}^{=}$ from now on, keeping $\mathbf{A}$ fixed. Next we define the following implicit function $\boldsymbol{\delta}(\boldsymbol{\lambda}) : \Delta^{m-1} \longrightarrow \mathbb{R}^m_+$ using the following equation system ($m-1$ independent equations since $\boldsymbol{\lambda}$ is in a $m-1$ dimensional space):

$$\boldsymbol{\lambda}^{=} = diag(\delta_1, \delta_2, ..., \delta_m)\boldsymbol{\lambda}$$

Lemma 1 $[\delta_i(\boldsymbol{\lambda})$ *is non-increasing in* i *and* $\boldsymbol{\lambda} \in \Delta^{m-1}] \iff [\boldsymbol{\lambda} \in \mathcal{R}_2(\mathbf{A})]$

Proof of Lemma 1. We proof in direction to the left by contradiction, suppose we have some $\tilde{\boldsymbol{\lambda}} \in \mathcal{R}_2(\mathbf{A}) \backslash \{\boldsymbol{\lambda}^{=}_{\mathbf{A}}\}$ with $\delta_i(\tilde{\boldsymbol{\lambda}}) > \delta_{i-1}(\tilde{\boldsymbol{\lambda}})$ for some $i, i-1 \in \mathcal{M}$. It has to be that

$$\sum_{j=i}^{m} \tilde{\lambda}_j < \sum_{j=i}^{m} \lambda^{=}_j \tag{14}$$

or else we already have a contradiction since the distribution $\boldsymbol{\lambda}^{=}$ has the most possible mass at the top subject to adding up to one, i.e., any $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}^{=}$ in $\mathcal{R}_2(\mathbf{A})$ will have the top component with less mass/frequency: $\lambda_m < \lambda^{=}_m$. Now, going back to inequality 14 define $\tilde{S} \equiv \sum\limits_{j=i}^{m} \tilde{\lambda}_j$ and $S^{=} = \sum\limits_{j=i}^{m} \lambda^{=}_j$ and multiply $\tilde{\boldsymbol{\lambda}}$ by $S^{=}/\tilde{S}$ to get

$$\tilde{\boldsymbol{\lambda}}' \equiv \tilde{\boldsymbol{\lambda}}(S^{=}/\tilde{S})$$

which, since wages are homogeneous, is equivalent for checking wage restrictions although it does not necessarily belong to Δ^{m-1}. Notice now that $\sum\limits_{j=i}^{m} \tilde{\lambda}_j = \sum\limits_{j=i}^{m} \lambda_j^{=}$ but $\tilde{\lambda}'_{i-1} > \lambda_{i-1}^{=}$ thus having that

$$\sum_{j=i-1}^{m} \tilde{\lambda}'_j > \sum_{j=i-1}^{m} \lambda_j^{=}$$

a contradiction to $\tilde{\boldsymbol{\lambda}}'$ satisfying $F_i(\mathbf{A}\tilde{\boldsymbol{\lambda}}') > F_{i-1}(\mathbf{A}\tilde{\boldsymbol{\lambda}}') \quad \forall i$, and therefore to $\tilde{\boldsymbol{\lambda}} \in \mathcal{R}_2(\mathbf{A})$. The other direction of the proof uses the same argument(s). *Q.E.D.*

It is now straightforward to prove that $\mathcal{R}_2(\mathbf{A})$ is convex:

Proof of Proposition 1 (continuation)
Take two elements of $\mathcal{R}_2(\mathbf{A})$, say $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. Then both $\delta_i(\boldsymbol{\lambda}_1)$ and $\delta_i(\boldsymbol{\lambda}_2)$ are non-increasing in i, as is their convex combination, so the convex combination of these two elements is in the same set again, by Lemma 1. *Q.E.D.*

B.2 Proof of Theorem 1

Following Stokey et al. (1989) we can designate this as a sequential problem (SP). If every finite sequential problem (i.e., $T < \infty$) has an (a unique) equilibrium, and the planner-dependant Value Functions $\tilde{V}_i$ have an upper bound then it follows by continuity that an (a unique) equilibrium exists for $T \longrightarrow \infty$.

Lemma 2 *Every finite SP has an equilibrium.*

Proof of Lemma 2. Start with backward induction at time T. $\boldsymbol{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T)$ exists because of compactness of the feasible set $\mathcal{F}()$ (and is unique if the h_i's are strictly concave because the objective function of the Social Planner, which is a linear combination of them, will be strictly concave). So at $T-1$, for *any* $\boldsymbol{\lambda} = \boldsymbol{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T) \in \mathcal{F}(\mathbf{A}_T, \mathbf{d})$, workers of type $i_0 < m$ will all choose the same amount of schooling for two reasons:

1) $P(\cdot)$ is strictly concave

2) Workers have rational expectations, i.e., they know $\boldsymbol{\Lambda}_T$, and thus are identical within skills.

Thus since $s_{ji} = s_i \ \forall j \in \mathcal{L}_{i,T-1}, \forall i \in \mathcal{M}$ we have that $\frac{1}{L_{T-1}} \int_{\mathcal{L}_{i,T-1}} B(\alpha_{ji}/\varsigma_{i+1}) dj = \lambda_i P(\alpha_i/\varsigma_{i+1})$ by the Law of Large Numbers, and so $\theta_i = P(\alpha_i/\varsigma_{i+1})$ in equilibrium $\forall i \in \mathcal{M}\backslash\{m\}$. So there is no aggregate uncertainty. Again, just before immigration occurs at $T-1$, all natives know the deterministic consequences of every (feasible) policy $\boldsymbol{\lambda}$, a fact that the Social Planner D_{T-1} takes into account for choosing $\mathbf{\Lambda}$ from, again a compact set, $\mathcal{F}(\mathbf{A}_{T-1}, \mathbf{d})$. One has to be careful at this point for showing uniqueness of $\mathbf{\Lambda}$ since D_{T-1}'s objective function does not directly inherit the strict concavity this time. However, it is strictly concave in this case as well (i.e., $\mathbf{\Lambda}$ is unique), this is shown in *Lemma 3* below. The process repeats itself over and over without any aggregate uncertainty ad infinitum (with uniqueness preserved in the corresponding case). *Q.E.D.*

Lemma 3 *$\mathbf{\Lambda}_t$ is unique for $t < T$ (given strict concavity of the h_i functions).*

Proof of Lemma 3. Note that if we show $\mathbf{\Lambda}_{T-1}$ is unique for all $\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1}$ then the Lemma follows directly.

Since $\mathbf{\Lambda}_{T-1}(\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1}) = \underset{\boldsymbol{\lambda}_{T-1} \in \mathcal{F}(\mathbf{A}_{T-1}, \mathbf{d})}{\arg\max} \left\{ \sum_{i=1}^{m} \eta_{i,T-1} \tilde{V}_{i,T-1}(\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1}, \boldsymbol{\lambda}_{T-1}) \right\}$, showing that the planner-dependant Value Functions $\tilde{V}_{iT-1}$ of the individuals are strictly concave in $\boldsymbol{\lambda}_{T-1}$ for any $(\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1})$ is sufficient. Consider these planner-dependant Value Functions from the Social Planner D_{T-1}'s perspective:

$$\begin{aligned}
\tilde{V}_{i,T-1}(\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1}, \boldsymbol{\lambda}_{T-1}) &= u(F_i(\mathbf{A}_{T-1}\boldsymbol{\lambda}_{T-1}) - \Theta_{i,T-1}(\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1})) \\
&\quad + \beta \left[P_i V_{i+1,T}(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})) + (1 - P_i) V_{i,T}(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})) \right] \\
\text{where} \quad V_{i,T}(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})) &= \tilde{V}_{i,T}\left(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}), \mathbf{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}))\right) \\
\text{and} \quad \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}) &= \mathbf{M}(\mathbf{\Theta}_{T-1}(\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1}))\boldsymbol{\lambda}_{T-1} \\
\text{and} \quad P_i &\equiv P(\Theta_{i,T-1}(\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1}))
\end{aligned}$$

where $P(\cdot)$ is the P-function defined in assumption 5. Notice D_{T-1} takes $\mathbf{A}_{T-1}, \boldsymbol{\eta}_{T-1}$ and $\mathbf{\Theta}_{T-1}$ as given, [23] and $u(\cdot)$ remains strictly concave in $\boldsymbol{\lambda}_{T-1}$. So all we need to show is the concavity of $V_{i,T}(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}))$ in λ_{T-1}. Notice that $V_{i,T}(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})) = \tilde{V}_{i,T}\left(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}), \mathbf{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}))\right) = \tilde{V}_{i,T}\left(\mathbf{A}_T, \mathbf{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}))\right)$, where on the RHS of the equality the second argument, $\boldsymbol{\eta}_T$, was dropped because at T, $\boldsymbol{\eta}_T$ has no effect on future utility (it is the last period, so the second argument of the function is meaningless).

That is, Lemma 3 follows from the concavity of $\tilde{V}_{i,T}\left(\mathbf{A}_T, \mathbf{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}))\right)$ with respect to $\boldsymbol{\lambda}_{T-1}$ proven in Claims 1 and 2, since the sum of a strictly concave and a concave function is strictly concave as well. This makes the maximizer unique (recall that the feasible set is convex and compact). *Q.E.D.*

Claim 1 $\mathbf{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}))$ *is concave in* $\boldsymbol{\lambda}_{T-1}$

[23]The former two because they are predetermined, the latter because he takes the citizens' schooling behavior as given (and state-dependent).

Proof of Claim 1. To see this consider:

$$\mathbf{\Lambda}_T\left(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})\right) = \underset{\mathbf{x}\in\mathcal{F}(\mathbf{A}_T,\mathbf{d})}{\arg\max}\left\{\sum_{i=1}^{m}\eta_{iT}(\boldsymbol{\lambda}_{T-1})\tilde{V}_{iT}(\mathbf{A}_T,\mathbf{x})\right\}$$

Stacking the functions $\left(\tilde{V}_{iT}(\mathbf{A}_T,\mathbf{x})\right)_{i=1}^{m}$ into the vector-valued function $\tilde{\mathbf{V}}_{iT}(\mathbf{A}_T,\mathbf{x})$ we can write

$$\begin{aligned}
\mathbf{\Lambda}_T\left(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})\right) &= \underset{\mathbf{x}\in\mathcal{F}(\mathbf{A}_T,\mathbf{d})}{\arg\max}\left\{\sum_{i=1}^{m}\eta_{iT}(\boldsymbol{\lambda}_{T-1})\tilde{V}_{iT}(\mathbf{A}_T,\mathbf{x})\right\} \\
&= \underset{\mathbf{x}\in\mathcal{F}(\mathbf{A}_T,\mathbf{d})}{\arg\max}\left\{\boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})^{\mathrm{T}}\,\tilde{\mathbf{V}}_{iT}(\mathbf{A}_T,\mathbf{x})\right\} \\
&= \underset{\mathbf{x}\in\mathcal{F}(\mathbf{A}_T,\mathbf{d})}{\arg\max}\left\{\left(\bar{\mathbf{M}}\boldsymbol{\lambda}_{T-1}\right)^{\mathrm{T}}\,\tilde{\mathbf{V}}_{iT}(\mathbf{A}_T,\mathbf{x})\right\}
\end{aligned}$$

where $\mathbf{v}^{\mathrm{T}}$ is the transpose of vector $\mathbf{v}$, and the last step follows from the fact that $\mathbf{\Theta}_{T-1}(\mathbf{A}_{T-1},\boldsymbol{\eta}_{T-1})$ is taken as a constant positive-definite transition matrix by D_{T-1}. And therefore

$$\mathbf{\Theta}_{T-1}(\mathbf{A}_{T-1},\boldsymbol{\eta}_{T-1})\boldsymbol{\lambda}_{T-1} = \bar{\mathbf{M}}\boldsymbol{\lambda}_{T-1} = \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})$$

for some $\bar{\mathbf{M}} \in \mathbb{R}_{+}^{m\times m}$.

We now show that $\mathbf{\Lambda}_T\left(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})\right)$ is linear in $\boldsymbol{\lambda}_{T-1}$. Let α be a positive constant, then:

$$\begin{aligned}
\alpha\,\mathbf{\Lambda}_T\left(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1})\right) &= \alpha\underset{\mathbf{x}\in\mathcal{F}(\mathbf{A}_T,\mathbf{d})}{\arg\max}\left\{\left(\bar{\mathbf{M}}\boldsymbol{\lambda}_{T-1}\right)^{\mathrm{T}}\,\tilde{\mathbf{V}}_{iT}(\mathbf{A}_T,\mathbf{x})\right\} \\
&= \underset{\mathbf{x}\in\mathcal{F}(\mathbf{A}_T,\mathbf{d})}{\arg\max}\left\{\alpha\left(\bar{\mathbf{M}}\boldsymbol{\lambda}_{T-1}\right)^{\mathrm{T}}\,\tilde{\mathbf{V}}_{iT}(\mathbf{A}_T,\mathbf{x})\right\} \\
&= \underset{\mathbf{x}\in\mathcal{F}(\mathbf{A}_T,\mathbf{d})}{\arg\max}\left\{\left(\bar{\mathbf{M}}\,\alpha\boldsymbol{\lambda}_{T-1}\right)^{\mathrm{T}}\,\tilde{\mathbf{V}}_{iT}(\mathbf{A}_T,\mathbf{x})\right\} \\
&= \mathbf{\Lambda}_T\left(\mathbf{A}_T, \boldsymbol{\eta}_T(\alpha\boldsymbol{\lambda}_{T-1})\right)
\end{aligned}$$

so the function is linear. Therefore the function is concave, which proves the Claim. *Q.E.D.*

Claim 2 $\tilde{V}_{i,T}\left(\mathbf{A}_T, \mathbf{\Lambda}_T(\mathbf{A}_T, \boldsymbol{\eta}_T(\boldsymbol{\lambda}_{T-1}))\right)$ *is concave in* $\boldsymbol{\lambda}_{T-1}$

Proof of Claim 2. Follows directly from the fact that both, $\tilde{V}_{i,T}\left(\mathbf{A}_T,\mathbf{x}\right)$ is concave in $\mathbf{x}$, and $\mathbf{\Lambda}_T(\mathbf{A}_T,\boldsymbol{\eta}_T(\mathbf{x}))$ is concave in $\mathbf{x}$. Therefore their composition is concave in $\mathbf{x}$ as well. (The composition of two concave functions is concave). *Q.E.D.*

B.3 Proof of Theorem 2

Following Stokey et al. (1989) we know that as long as $[0, F_i(\mathbf{A}\mathbf{\Lambda}(\mathbf{A}, \boldsymbol{\eta}))] \neq \emptyset$ for all $\mathbf{A}, \boldsymbol{\eta}$ (which is the case with $\mathcal{F}(\mathbf{A}, \mathbf{d})$ non-empty) and a limit for all feasible sequences exists (shown previously by the boundedness using $\bar{\beta}$), the sequential problem is well defined and V_i satisfies the usual Functional Equation employed in Dynamic Programming (Theorem 4.2 in Stokey et al.). Since this is an equilibrium, it must be that, by construction, $\mathbf{\Lambda}$ and $\mathbf{\Theta}$ satisfy the functional equations as well. $\mathbf{\Lambda}$ is continuous by the convexity and compactness of the opportunity set $\mathcal{F}(\mathbf{A}, \mathbf{d})$ and $\mathbf{\Theta}$ is continuous by the continuity of the maximizer $s_i^*(\mathbf{A}, \boldsymbol{\eta})$ for all i. *Q.E.D.*

B.4 Proof of Theorem 3

We proceed in three steps. First we show that by a Fixed-Point argument it is possible to show the existence of such a steady state. Then we show that the two assumed conditions in step one are satisfied.

B.4.1 Existence of a steady state

In a steady state we have that the law of motion satisfies:

$$\boldsymbol{\eta}_{SS} = \mathbf{M}(\boldsymbol{\theta}_{SS})\boldsymbol{\lambda}_{SS}$$

Since this is an equilibrium, we also have that

$$\begin{aligned} \boldsymbol{\lambda}_{SS} &= \mathbf{\Lambda}(\mathbf{A}_t, \boldsymbol{\eta}_{SS}) \ \forall t \\ \boldsymbol{\theta}_{SS} &= \mathbf{\Theta}(\mathbf{A}_t, \boldsymbol{\eta}_{SS}) \ \forall t \end{aligned}$$

Thus it must be that both policy functions $\mathbf{\Lambda}, \mathbf{\Theta}$ are time-invariant. Assume this is true for now (it will be proven in the other two steps). Then it follows that

$$\boldsymbol{\eta}_{SS} = \mathbf{M}(\mathbf{\Theta}(\boldsymbol{\eta}_{SS}))\mathbf{\Lambda}(\boldsymbol{\eta}_{SS})$$

Since $\boldsymbol{\eta} \in \Delta^{m-1}$ (a convex and compact set), and all functions are continuous, it follows by Brouwer's Theorem that such a vector $(\boldsymbol{\eta}_{SS})$ exists.

B.4.2 The constancy of $\mathbf{\Theta}(\mathbf{A}_t, \boldsymbol{\eta}_{SS})$

Since in equilibrium $\Theta_i(\mathbf{A}_t, \boldsymbol{\eta}_{SS}) = P(s_i^*/z_{i+1}) = P(\alpha_i^*(\mathbf{A}_t, \boldsymbol{\eta}_{SS})/\varsigma_i)$ it is enough to show the constancy of α_i^*.

It can be shown with some algebra that in a such a steady state, the Value functions can be represented as follows:

$$V_i(\mathbf{A}_t, \boldsymbol{\eta}_{SS}) = \sum_{h=i}^{m} \Gamma_h u(w_{h,t}) \tag{15}$$

Where the Γ_h's are some time-invariant coefficients and $w_{h,t}$ is the wage of type h at time t.

Hence the First Order Condition for i's choice of α can be rewritten like:

$$\Gamma'_0 = \sum_{h=i}^{m} \Gamma'_h \frac{u(w_{h,t})}{u'(w_{i,t})w_{i,t}}$$

Where the Γ's again are some time-invariant coefficients.

It can be shown that, given Assumption 1, the RHS of this equation is actually time-invariant, thus proving the constancy of α (for all i's) and hence for $\boldsymbol{\Theta}$.

B.4.3 The constancy of $\boldsymbol{\Lambda}(\mathbf{A}_t, \boldsymbol{\eta}_{SS})$

Recall that the social planner chooses $\boldsymbol{\lambda}$ to maximize his Objective Function:

$$\sum_{i \in \mathcal{M}} \eta_{i,SS} V_i(\mathbf{A}_t, \boldsymbol{\eta}_{SS})$$

From equation 15 we know that this Objective function can be represented as follows:

$$\sum_{i \in \mathcal{M}} \sum_{h=i}^{m} \Gamma_h u(w_{h,t})$$

Moreover by Assumption 1 and the fact that in such a steady state the wage of type h can be represented as $w_{h,t} = w_{h,0} g^t$ we can further simplify the Objective function to get:

$$u(g^t) \sum \Gamma''$$

Where the Γ''s are again time-invariant. Thus showing that the optimal $\boldsymbol{\lambda}$ is independent of t. *Q.E.D*

B.5 Proof of Theorem 4

B.5.1 Nonexistence of a steady state

Let us proof the nonexistence of a steady state by contradiction,

Suppose there was a steady state η_{SS}, then even if λ wasn't stationary, we would have $\eta_{SS} = M(\theta_t^*)\lambda_t^*$ thus necessarily having that η_{SS} first order dominates λ_t^* for all t. In other words, λ_m would be bounded and so would $\sum_{i=j}^{m} \lambda_i$ for all $j > 1$, thus having that wages for the high skill would eventually start increasing without bounds, implying an increasing inequality under the same native electorate η_{SS}, which directly contradicts the Immigration Policy characterized in Theorem 3.

B.5.2 Inequality

Consider now inequality. It is enough if we show that $V_m > V_{m-1}$ since then the rest of the inequalities follow.

Let us again proceed by contradiction supposing that $V_m = V_{m-1}$, then it would have to be the case that $P(s_{m-1}^*) = \theta_{m-1} = 0$ and that the excess demand of the top skill (the minimal amount of top skilled workers needed to keep wages of types m and $m-1$ equal) is satisfied with top-skilled immigration.

But then we have that an infinitesimal increase in immigration-induced inequality would make everyone weakly better off (and some strictly), a Pareto improvement that is always done by the equilibrium Immigration Policy Λ. Again, this is due to the infinite returns of the P function at the origin.

B.5.3 $\iota_{m,t}$ becomes positive in finite time

Let us now show that schooling at the next-to-top skill goes to zero as time increases. Since, for a suitably small discount factor, $\lambda_{m,t}$ increases with time, it must be that $\iota_m > 0$ eventually.

If a unique equilibrium exists, it must be that the Value functions V_i are well-defined (they exist and are finite). Therefore we can construct a new Value function, say

$$\hat{V}_{m,t} = \frac{u(\hat{w}_{i,0})}{1 - \beta u(\hat{w}_{m,0})} u(\hat{g}_m)^t$$

for some $\hat{w}_{m,0}, \hat{g}_m \in \mathbb{R}_{++}$ such that $\hat{V}_{m,t} \geq V_{m,t}$. Along the same lines, we know $w_{1,t} \geq d_{1,t}$ so we can construct a Value function like

$$\hat{V}_{1,t} = \frac{u(d_{1,0})}{1 - \beta u(g_d)} u(g_d)^t$$

which satisfies $\hat{V}_{1,t} \leq V_{1,t}$ (recall $g_d > 1$). Therefore we have that

$$\Delta V_{m,t} \leq \Delta \hat{V}_{m,t}$$

Consider now the First Order Condition of $i = m - 1$ (and bear in mind that $s_{i,t} = \alpha_{i,t} w_{i,t}$ and $z_{i+1,t} = \varsigma_i w_{i,t}$):

$$w_{i,t} u'([1 - \alpha_{i,t}] w_{i,t}) = \frac{\beta}{\varsigma_i} P' \left(\frac{\alpha_{i,t}}{\varsigma_i} \right) \Delta V_{i+1,t+1}$$

Therefore, using our bounding Value functions we know that our bounding $\hat{\alpha}$ will be at least as big as the original α:

$$w_{i,t} u'([1 - \hat{\alpha}_{i,t}] w_{i,t}) = \frac{\beta}{\varsigma_i} P' \left(\frac{\hat{\alpha}_{i,t}}{\varsigma_i} \right) \Delta \hat{V}_{i+1,t+1}$$

It can then be shown that $\hat{\alpha}$ will go to zero when t goes to infinity as long as the following is true:

$$\hat{g}_d > \frac{u(\hat{g}_m)}{u'(g_d)}$$

That is, as long as

$$\left(\frac{g_d}{\hat{g}_m} \right)^{1-b_2} > \frac{1}{1 - b_2}$$

Where b_2 is the coefficient of relative risk aversion from our Instant Utility function (see assumption 1). This inequality is trivially satisfied for $b_2 > 1$. *Q.E.D.*

C Appendix: Source Tables for Figures in Section 3.1

C.1 USA

Table 3: USA 1940

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
None or preschool	1.29	6.58	5.29	6.823252
Grade 1, 2, 3, or 4	6.7	12.76	6.06	5.742843
Grade 5, 6, 7, or 8	40.86	49.96	9.1	3.895809
Grade 9	6.97	4.31	-2.66	0.674101
Grade 10	8.22	5.31	-2.91	-1.928232
Grade 11	4.53	2.25	-2.28	-3.584191
Grade 12	18.59	11.71	-6.88	-4.457664
1 to 3 years of college	7.04	3.33	-3.71	-4.06044
4+ years of college	5.8	3.78	-2.02	-3.115477

Table 4: USA 1950

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
None or preschool	0.93	3.13	2.20	3.391199
Grade 1, 2, 3, or 4	4.88	8.03	3.14	3.285291
Grade 5, 6, 7, or 8	28.21	35.08	6.87	2.583785
Grade 9	6.82	5.31	-1.51	0.618434
Grade 10	8.43	6.56	-1.88	-1.135899
Grade 11	6.17	4.37	-1.80	-2.268568
Grade 12	28.05	23.62	-4.42	-2.740978
1 to 3 years of college	9.08	7.20	-1.88	-2.280249
4+ years of college	7.43	6.71	-0.72	-1.453014

Table 5: USA 1960

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
None or preschool	0.66	3.19	2.53	4.201203
Grade 1, 2, 3, or 4	2.81	7.25	4.44	4.060947
Grade 5, 6, 7, or 8	19.57	27.35	7.78	3.085089
Grade 9	6.89	4.89	-2	0.627556
Grade 10	8.59	6.67	-1.92	-1.610272
Grade 11	6.69	5.87	-0.82	-3.240793
Grade 12	34.62	24.1	-10.52	-4.031267
1 to 3 years of college	10.43	10.53	0.1	-2.53856
4+ years of college	9.75	10.15	0.4	-0.563902

Table 6: USA 1970

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
None or preschool	0.73	2.57	1.84	4.141774
Grade 1, 2, 3, or 4	1.32	5.75	4.43	4.988774
Grade 5, 6, 7, or 8	10.81	22.77	11.96	4.684887
Grade 9	5.65	4.93	-0.72	1.799839
Grade 10	7.77	5.91	-1.86	-1.459086
Grade 11	6.55	4.75	-1.8	-4.144525
Grade 12	40.25	26.23	-14.02	-5.509572
1 to 3 years of college	12.65	11.88	-0.77	-3.635059
4+ years of college	14.26	15.21	0.95	-0.857031

Table 7: USA 1980

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
None or preschool	0.41	2.32	1.91	4.209877
Grade 1, 2, 3, or 4	0.5	5.39	4.89	5.32388
Grade 5, 6, 7, or 8	4.73	16.4	11.67	5.287945
Grade 9	3.3	3.99	0.69	2.735194
Grade 10	4.44	3.88	-0.56	-0.510225
Grade 11	4.25	3.75	-0.5	-3.64676
Grade 12	40.2	25.81	-14.39	-5.897747
1 to 3 years of college	20.59	16.79	-3.8	-4.913141
4+ years of college	21.59	21.68	0.09	-2.589023

Table 8: USA 1990

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
No school completed	0.44	4.25	3.81	4.867264
1st-4th grade	0.16	4.21	4.05	5.301604
5th-8th grade	1.79	11.83	10.04	5.207312
9th grade	1.82	3.66	1.84	3.429954
10th grade	2.97	2.86	-0.11	1.231439
11th grade	3.03	2.46	-0.57	-0.921298
12th grade, no diploma	2.82	5.13	2.31	-3.232046
High school graduate, or GED	32.03	19.17	-12.86	-5.728941
Some college, no degree	22.66	15.32	-7.34	-5.669099
Associate degree	8.43	7.09	-1.34	-3.875164
Bachelors degree	16.46	14.72	-1.74	-2.005229
Masters degree	5.08	5.78	0.7	-0.449808
Professional degree	1.75	2.16	0.41	0.533203
Doctorate degree	0.54	1.29	0.75	1.260809

Table 9: USA 2000

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
No school completed	0.46	4.27	3.81	4.305049
1st-4th grade	0.1	2.28	2.18	5.089005
5th-8th grade	1.27	12.63	11.36	5.625435
9th grade	1.5	4.8	3.3	4.212314
10th grade	2.44	2.72	0.28	2.014896
11th grade	2.76	2.54	-0.22	-0.257719
12th grade, no diploma	3.23	5.96	2.73	-2.763882
High school graduate, or GED	30.35	19.27	-11.08	-5.643083
Some college, no degree	24.16	14.86	-9.3	-6.287871
Associate degree	8.34	5.79	-2.55	-4.809254
Bachelors degree	18.16	15.02	-3.14	-2.824306
Masters degree	5.01	6.05	1.04	-0.82047
Professional degree	1.7	2.37	0.67	0.556959
Doctorate degree	0.52	1.43	0.91	1.592926

C.2 Canada

Table 10: Canada 1971

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
No schooling	0.7	1.37	0.67	1.391992
Below grade 5	2.07	4.35	2.28	-0.512506
Grades 5-8	26.44	25.48	-0.96	-2.778
Grades 9-10 (1971)	26.41	15.99	-10.42	-4.369234
Grade 11 (1971)	13.03	8.1	-4.93	-3.34195
Grade 12 (1971)	15.87	15.94	0.07	-0.777274
Grade 13 (1971)	3.07	9.74	6.67	1.449642
University, years 1-2, no degree	4.36	5.14	0.78	1.887284
University, years 3-4, no degree	1.36	2.71	1.35	1.694317
University, years 5+, no degree	0.22	0.76	0.54	1.475764
University, years 3-4, degree	3.2	4.19	0.99	1.664489
University, years 5+, degree	3.26	6.23	2.97	2.225477

Table 11: Canada 1981

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
Below grade 5	1.23	3.67	2.44	1.614216
Grades 5-8	10.07	11.03	0.96	-0.873963
Grades 9-13	22.64	16.3	-6.34	-2.94925
High school graduation certificate	15.35	10.06	-5.29	-3.281772
Trades certificate or diploma	3.69	3.83	0.14	-2.237029
Non-university without trades or college certificate or diploma	6.26	5.13	-1.13	-1.184638
Non-university with trades certificate or diploma	8.77	9.83	1.06	-0.3057
Non-university with college certificate or diploma	10.15	9.89	-0.26	0.246004
University, no certificate, diploma or degree	3.73	4.06	0.33	0.999542
University or college certificate or diploma	5.74	7.25	1.51	2.230979
Bachelor or first professional degree	12.37	17.09	4.72	3.881611

Table 12: Canada 1991

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
Below grade 5	0.6	2.15	1.55	1.476947
Grades 5-8	4.18	6.12	1.94	-0.261591
Grades 9-13	19.49	14.74	-4.75	-1.963603
High school graduation certificate	16.89	13.61	-3.28	-2.491766
Trades certificate or diploma	4.76	3.18	-1.58	-2.101959
Non-university without trades or college certificate or diploma	7.14	6.88	-0.26	-1.444174
Non-university with trades certificate or diploma	8.57	7.59	-0.98	-0.907426
Non-university with college certificate or diploma	12.88	11.4	-1.48	-0.288642
University, no certificate, diploma or degree	3.88	5.04	1.16	0.578965
University or college certificate or diploma	6.15	7.55	1.4	1.2665
Bachelor or first professional degree	11.39	13.79	2.4	1.635588
Certificate or diploma above bachelor level	1.64	2.29	0.65	1.614603
Master's degree	2.17	4.53	2.36	1.514125
Doctoral degree	0.24	0.9	0.66	1.162431

Table 13: Canada 2001

Educational Level	Native Distribution	Immigrant Distribution	Phi	H-P Phi
Below grade 5	0.64	1.36	0.72	1.120468
Grades 5-8	1.75	3.47	1.72	0.194037
Grades 9-13	15.32	13.49	-1.83	-0.932627
High school graduation certificate	14.65	12.2	-2.45	-1.696779
Trades certificate or diploma	3.82	2.55	-1.27	-1.984356
Non-university without trades or college certificate or diploma	7.7	5.54	-2.16	-2.057908
Non-university with trades certificate or diploma	8.94	7.79	-1.15	-1.822807
Non-university with college certificate or diploma	17.18	13.49	-3.69	-1.235469
University, no certificate, diploma or degree	3.53	4.5	0.97	0.08409
University or college certificate or diploma	6.9	8.6	1.7	1.288592
Bachelor or first professional degree	14.6	18.17	3.57	1.973713
Certificate or diploma above bachelor level	1.88	2.28	0.4	1.940832
Master's degree	2.77	5.48	2.71	1.789473
Doctoral degree	0.31	1.08	0.77	1.348742

References

[1] Acemoglu, D. (2003): "Patterns of Skill Premia," *Review of Economic Studies*, 70, 199-230.

[2] Acemoglu, D., and J. A. Robinson (2000): "Why Did the West Extend the Franchise? Democracy, Inequality and Growth in Historical Perspective," *Quarterly Journal of Economics*, 115, 1167-1199.

[3] Adda, J., and R. Cooper (2003): *Dynamic Economics*, Cambridge, Massachusetts, MIT Press.

[4] Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutstrom (2008): "Eliciting Risk and Time Preferences," *Econometrica*, 76, 583-618.

[5] Barro, R. I., and X. Sala-i-Martin (2001): *Economic Growth*, Cambridge, Massachusetts, MIT Press.

[6] Bartel, A. P., and F. R. Lichtenberg (1987): "The Comparative Advantage of Educated Workers in Implementing New Technology," *Review of Economics and Statistics*, 69, 1-11.

[7] Benhabib, J. (1996): "On the Political Economy of Immigration," *European Economic Review*, 40, 1737-1743.

[8] Borjas, G. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531-553.

[9] Borjas, G. (2003): "The Labor Demand is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, 118, 1335-1374.

[10] Borjas, G., and V. Ramey (1994): "Time-Series Evidence on the Sources of Trends in Wage Inequality," *American Economic Review*, 84, 10-16.

[11] Champlin, D., and E. Hake (2006): "Immigration as Industrial Strategy in American Meatpacking," *Review of Political Economy*, 18, 49-69.

[12] Chiquiar D., and G. Hanson (2005): "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal Of Political Economy*, 113, 239-281.

[13] Epstein, L. (1974): "Some Economic Effects of Immigration: A General Equilibrium Analysis," *Canadian Journal of Economics*, 7, 174-190.

[14] Goldin, C. (1994): "The Political Economy of Immigration Restriction in the U.S., 1890 to 1921," in *The Regulated Economy: A Historical Approach to Political Economy*, ed. by C. Goldin and G. Lebecap. Chicago and London, University of Chicago Press, 223-257.

[15] Gomez-Ruano G. (2007): "The Political Economy of Immigration: Limits of the Median Voter Model," Working Paper, Boston University.

[16] Grossman, G. M., and E. Helpman (1996): "Electoral Competition and Special Interest Politics," *Review of Economic Studies*, 63, 265-286.

[17] Hamermesh, D. S. (1993): *Labor Demand*, Princeton, Princeton University Press.

[18] Hanson, G. (2005): *Why Does Immigration Divide America? Public Finance and Political Opposition to Open Borders*, Washington D.C., Institute for International Economics.

[19] Hanson, G., and A. Spilimbergo (2001): "Political Economy, Sectoral Shocks, and Border Enforcement," *Canadian Journal of Economics*, 34, 612-638.

[20] Hassler, J., P. Krusell, K. Storesletten and F. Zilibotti (2005): "The Dynamics of Government," *Journal of Monetary Economics*, 52, 1331-1358.

[21] Hillman, A. L., and A. Weiss (1999): "A Theory of Permissible Illegal Immigration," *European Journal of Political Economy*, 15, 585-604.

[22] Hodrick R. J., and E. C. Prescott (1997): "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, 29, 1-16.

[23] Judd, K. L. (1998): *Numerical Methods in Economics*, Cambridge, Massachusetts, MIT Press.

[24] Katz, L., and K. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35-78.

[25] Krusell, P., L. Ohanian, J. Rios-Rull and J. Violante (2000): "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 68, 1029-1053.

[26] Lang, K. (2007): *Poverty and Discrimination*, Princeton, Princeton University Press.

[27] Lindbeck, A., and J. Weibull (1987): "Balanced-Budget Redistribution as the Outcome of Political Competition," *Public Choice*, 52, 273-297.

[28] Loury, G. C. (1976): "Essays in the Theory of the Distribution of Income," *Ph.D. Thesis*, Massachusetts Institute of Technology, 1976.

[29] Loury, G. C. (1981): "Intergenerational Transfers and the Distribution of Earnings," *Econometrica*, 49, 843-867.

[30] Minnesota Population Center (2008): *Integrated Public Use Microdata Series — International: Version 4.0.*, Minneapolis, University of Minnesota.

[31] Mas-Collel, A. M., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*, Oxford, Oxford University Press.

[32] Miranda, M. J., and P. L. Fackler (2002): *Applied Computational Economics and Finance*, Massachusetts, USA, MIT Press.

[33] Mookherjee, D., and D. Ray (2002): "Is Equality Stable?," *American Economic Review*, 92, 253-259.

[34] OECD (2001): *Trends in International Migration, Annual Report*, 2001 Edition, Paris, France.

[35] Ortega, F. (2005): "Immigration Quotas and Skill-Upgrading," *Journal of Public Economics*, 89, 1841-1863.

[36] ORourke, K., and R. Sinnott (2006): "The Determinants of Individual Attitudes towards Immigration," *European Journal of Political Economy*, 22, 838-861.

[37] Ruggles, S., M. Sobek, T. Alexander, C. Fitch, R. Goeken, P. Hall, M. King and C. Ronnander (2004): *Integrated Public Use Microdata Series*: Version 3.0, Minneapolis MN, Minnesota Population Center.

[38] Solon, G. (1992): "Intergenerational Mobility in the United States," *American Economic Review*, 82, 393-408.

[39] Stokey, N. L., R. E. Jr Lucas and E. C. Prescott (1989): *Recursive Methods in Economic Dynamics*, Cambridge and London, Harvard University Press.